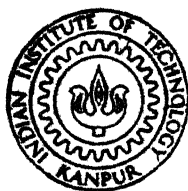# ESTIMATION AND PREDICTION IN LONGITUDINAL STUDIES

*by*

MARKAND PUSHPAK OZA

DEPARTMENT OF MATHEMATICS

## INDIAN INSTITUTE OF TECHNOLOGY KANPUR

OCTOBER, 1986

# ESTIMATION AND PREDICTION IN LONGITUDINAL STUDIES

*A Thesis Submitted*
In Partial Fulfilment of the Requirements
for the degree of
DOCTOR OF PHILOSOPHY

*by*

MARKAND PUSHPAK OZA

*to the*

DEPARTMENT OF MATHEMATICS
# INDIAN INSTITUTE OF TECHNOLOGY KANPUR
OCTOBER, 1986

Th
519·54
0Z1 e

CERTIFICATE

This is to certify that the work done in this thesis, entitled "ESTIMATION AND PREDICTION IN LONGITUDINAL STUDIES", has been carried out by Markand Pushpak Oza under my supervision and guidance. To the best of my knowledge, this work has not been submitted elsewhere for award of any diploma or degree.

Dr. G.K. Shukla
Assistant Professor
Department of Mathematics
Indian Institute of Technology
Kanpur, India

# ACKNOWLEDGEMENTS

# LIST OF TABLES

SYNOPSIS

of the

Ph. D. Dissertation

on

ESTIMATION AND PREDICTION IN LONGITUDINAL STUDIES

by

Markand Pushpak Oza
Department of Mathematics
Indian Institute of Technology, Kanpur (India)

September, 1986

Very often one comes across studies where observations are made repeatedly on the same experimental unit over a period of time. Such observations may be available for a number of experimental units. These studies are generally known as longitudinal studies. There are certain advantages of such studies over cross-sectional studies, for details see Ware (The American Statistician, 1985, 39, 95-101). The emphasis of longitudinal studies is on the explanation of changes within units. The purpose of the present study is to fit an appropriate model that can summarise the information contained in the data. Estimation of model parameters is, of course, an important aspect in the analysis. But the main interest of the investigators is usually in prediction rather than in estimation, because prediction involves the response which investigators actually measure. With this in mind, considerable emphasis is laid in this thesis on prediction problems.

The thesis consists of five chapters. A chapterwise summary is as follows :

Chapter 1 is introductory in nature. It explains the purpose of this study. Relevant literature on growth curves and repeated measurement studies is briefly reviewed. Expectation and maximization (EM) algorithm for obtaining the maximum likelihood (ML) estimates, an iterative technique that will often be used during the course of the thesis, is also explained.

In Chapter 2, we consider the choice of optimum degree of polynomial and problem of prediction under uncertainty about the degree of the polynomial in random coefficient regression models. Two strategies for selection of covariates for covariance adjusted estimators are also discussed. Some of the results of this chapter have been published in Oza and Shukla (Commun. Statist. -Theor. Meth., 1986, 15(5), 1653-1664).

A generalised multivariate growth curve model, which allows different design matrix and different dimension of regression parameters for each characteristic, is consided in Chapter 3. The random regression coefficients for each characteristic may also depend upon different set of auxiliary variables. This generalises the model of Reinsel (J. Am. Stat. Assoc., 1984, 79, 406-414) in many ways. The ML procedure is adopted for estimation of the parameters. The problems of prediction and derivation of its asymptotic mean square error (MSE) are considered. A Monte Carlo simulation study is carried

out to compare the performance of various predictors including an empirical Bayes type predictor. A possible application in preharvest forecasting is considered in great details.

Anderson (J. Am. Stat. Assoc., 1978, 73, 371-378) has considered a multivariate autoregressive process for repeated series. From the practical point of view there is a need to generalise this model to a two-stage model. This is done in Chapter 4. In the first stage the usual regression model is considered and in the second stage the regression parameters of the first stage are assumed to follow a first order auto-regressive process. An EM-algorithm for the computation of ML estimates of the model parameters is derived and prediction under this model is discussed. Some simple estimators, as an alternative to ML estimators, are also considered. A Monte Carlo simulation study is performed to compare these methods of estimation. Some possible applications of such two-stage models are discussed.

A general mixed model with autocorrelated errors, which is similar to one of the models considered by Swamy (Statistical Inference in Random Coefficient Regression Models, 1971, Springer-Verlag, Berlin), is considered in Chapter 5. The ML estimators are obtained using EM-algorithm. Prediction using these estimated values is considered. An expression for its approximate MSE is derived. A Monte Carlo simulation study is performed to compare the behaviour of estimators and predictors based on the ML method and some other methods.

# CHAPTER I

## INTRODUCTION

## 1.1 INTRODUCTION

It is very difficult to give a universally acceptable definition of the term longitudinal studies. However, there are certain features which are common to all longitudinal studies. In these studies measurements on the same characteristic are obtained from the same experimental units at two or more occasions. Data on several units may be available and units may be assumed to be selected randomly from a population of units.

The purpose of the longitudinal studies is usually to explain the changes taking place in an experimental unit between occasions and attribute these changes to certain background variables. There are certain advantages of longitudinal studies over cross-sectional studies. The response may be affected by uncontrolled variables and in such situations the longitudinal studies generally produce more efficient estimates of within unit changes than the cross-sectional studies. When the population is of limited size, as in the case of a hospital, there is no choice but to depend on longitudinal studies. In a recent paper Ware (1985) has given a good description of application of linear models in longitudinal studies. However, there are some difficulties in

longitudinal studies, such as the observations may be correlated, the covariates may vary with time, observations may be missing at some points etc.. This makes the usual statistical analysis, developed for independent observations, unsuitable for the present purposes. If the main purpose is to study the variability over units then the longitudinal studies may not be very efficient. The longitudinal studies are also time consuming.

Some work on hybrid of longitudinal and cross-sectional studies, called mixed studies, has also been reported in literature. One of the first such studies is due to Rao and Rao (1966) and they termed it as linked cross-sectional study. Such studies are less time consuming and provide quick estimates of growth rate than that by a purely longitudinal study. For more details on method of analysis one may refer Rao and Rao (1966), Woolson et al. (1978) and Woolson and Leeper (1980). For a review of the literature on the topic see Dielman (1983).

In longitudinal studies, the observations on an experimental unit are ordered in time and hence form a time series. But for a time series analysis to be valid one needs to have many observations - certainly many times more than what we have in a typical longitudinal study. Therefore, there is a need to develop separate techniques for analysis of the longitudinal data. A number of specialised methods are used in the analysis of the longitudinal data under the title of growth curve models.

The analysis simplifies considerably when observations are made on each experimental unit at a set of fixed time points.

For the purpose of analysis, the usual method of fitting a model is followed. The reason to fit a model is that it can summarise the information contained in the data. Estimation of model parameters is, of course, an important aspect in the analysis. But since early-seventies the emphasis has shifted from estimation to prediction. This change has taken place due to recognition of the fact that the main interest of the investigators is usually in prediction rather than estimation, because prediction involves the response which investigators actually measure. With this in mind, considerable emphasis is laid in this thesis on prediction problems.

The structure of the thesis is as follows : In section 1.2 of this chapter we review the literature on growth and repeated measurement models. Motivation for the present work is discussed in section 1.3. Expectation and maximization (EM) algorithm for obtaining the maximum likelihood (ML) estimates is explained in section 1.4. In Chapter II, we consider the choice of optimum degree of polynomial for prediction under uncertainty about the degree of the polynomial in random coefficient regression (RCR) models. Two strategies for selection of covariates for covariance adjusted estimators are also proposed. Estimation and prediction in a multivariate RCR model

when designs are different for different characteristics is considered in Chapter III. The dimension of regression parameters and set of covariates for each characteristic may also be different. A two-stage autoregressive model for repeated measurements is proposed and studied in Chapter IV. A general mixed model with autocorrelated errors is considered in Chapter V.

## 1.2 REVIEW OF THE LITERATURE

1.2.1 <u>Model and estimation</u> :   In growth curve model studied frequently in literature we have N independent individuals, each providing an observable p component random vector $\underset{\sim}{y}$ having multivariate normal distribution with dispersion matrix $\Omega$ and unknown mean $X\underset{\sim}{\beta}$. Here X is a known pxq design matrix of rank $q \leq p$ and $\underset{\sim}{\beta}$ is an unknown q component vector of regression parameters. This is called a one-sample growth curve model. Wishart (1938) is credited for using such growth curve models for the first time. His analysis begins by replacing large number of observations on each individual by a few coefficients of fitted orthogonal polynomial. The means of growth rate and its change are compared by univariate analysis of variance. Box (1950) suggests the use of analysis of variance technique to the first difference of successive observations when the assumption of uniform covariance structure is valid. He hints at the multivariate techniques when the assumption of uniform covariance matrix is not valid. Rao (1958) introduces the concept of time metameter i.e. of transforming the variables so

that the data are linear with respect to the transformed variables. The comparison of growth parameters is then made by the use of analysis of variance or Wilks' $\Lambda$-criterion. Rao (1959) has obtained the least square estimator of $\underset{\sim}{\beta}$ when dispersion matrix $\Omega$ is unknown but estimate of it is available. He also considers the problems of inference regarding $\underset{\sim}{\beta}$ and setting confidence bands for a linear function of $\underset{\sim}{\beta}$.

A k-sample version of this model is considered by Rao (1961). Glesser and Olkin (1966) consider the inference problems for such models in detail. They obtain simultaneous maximum likelihood estimators of regression parameters of k-samples i.e. $\underset{\sim}{\beta}^{(1)}, \underset{\sim}{\beta}^{(2)}, \ldots, \underset{\sim}{\beta}^{(k)}$, and the common dispersion matrix $\Omega$. They also obtain the distribution of these estimators. They derive the likelihood ratio test for testing equality of $\underset{\sim}{\beta}^{(1)}, \underset{\sim}{\beta}^{(2)}, \ldots, \underset{\sim}{\beta}^{(k)}$ and distributions of likelihood ratio test statistics.

Potthoff and Roy (1964) introduce a polynomial growth curve model which is a generalisation of the multivariate analysis of variance (MANOVA) model. The model is of the form $E(Y) = X\theta W$ where $Y$ is $p \times N$ matrix of observable random variables whose columns are independently distributed with unknown dispersion matrix $\Omega$, $W$ is a $k \times N$ matrix across individuals and $\theta$ is a $q \times k$ matrix of unknown parameters. Their method of analysis is a multivariate analogue of weighted least square method. The most severe limitation of this approach is the absence of a

well defined matrix G occuring in the transformation to the
usual MANOVA model. For any choice of symmetric matrix G such
that $G^{-1}$ and $(X'G^{-1}X)^{-1}$ exist, the estimator of $\theta$ is unbiased
and is the most efficient when $G = \Omega$. Khatri (1966) tackles
the problem of arbitrary G in Potthoff-Roy analysis and obtains
maximum likelihood estimator of $\theta$ when $\Omega$ is unknown. He also
obtains likelihood ratio test of $H_o$ : $C\theta D = 0$, a linear
combination of some estimable function of $\theta$, and its confidence
bands.

Some work on incomplete growth data is also reported
in the literature. Trawinski and Bargman (1964) develop
procedures for estimation and tests of hypotheses in one sample
growth curve models. Klienbaum (1973) considers the case of
missing observations in the generalised growth curve model of
Potthoff and Roy (1964). Analysis of incomplete multiresponse
designs is considered by Srivastava (1966,1968) and Roy et al.
(1971). All of them consider fixed effect models with general
dispersion matrix and view it as a situation in the set up of
MANOVA model. A review of MANOVA of repeated measurements is
given by Timm (1980).

Some amount of variability in observable random
vectors $\underset{\sim}{y}_i$ may be due to random variation in the individuals
themselves. The random coefficient regression (RCR) model
proposed by Rao (1965) takes into account the variability due
to individuals. At the first stage, the model has the following
form;

$$\underset{\sim}{y}_i = X\underset{\sim}{\beta}_i + \underset{\sim}{e}_i$$

with

$$E(\underset{\sim}{e}_i|\underset{\sim}{\beta}_i) = \underset{\sim}{0} \; ; \; var(\underset{\sim}{e}_i|\underset{\sim}{\beta}_i) \; \sigma^2 I.$$

This is known as the conditional model considering $\underset{\sim}{\beta}_i$'s as fixed. If we assume $\underset{\sim}{\beta}_i$'s to be random variables with

$$E(\underset{\sim}{\beta}_i) = \underset{\sim}{\beta} \; ; \; var(\underset{\sim}{\beta}_i) = F \; ; \; cov(\underset{\sim}{\beta}_i, \underset{\sim}{\beta}_j) = 0, \; i \neq j \; ;$$

$$cov(\underset{\sim}{\beta}_i, \underset{\sim}{e}_j) = 0 \text{ for all } i \text{ and } j,$$

then we get an RCR model. This model imposes a structure on dispersion matrix that $Q = XFX' + \sigma^2 I$. Rao (1965) obtains least square estimators of $\underset{\sim}{\beta}_i$ and $\underset{\sim}{\beta}$. If an unbiased estimator of $Q$ which has a Wishart distribution independent of $\underset{\sim}{y}_i$'s exists, then he derives the likelihood ratio test for testing appropriateness of the model and the structure of $Q$. Fisk (1967) also considers the problem of estimation in RCR models. Properties of various estimators of the population regression parameters are also studied. A thorough review of the RCR model has been given by Spjotvoll (1977); he mentions many possible applications of the model and certain problems for future research.

Rao (1966) considers covariance adjustment in Potthoff-Roy model of growth curves. Least square estimation and inference on $\underset{\sim}{\beta}$ for growth curve models when $Q$ has various structures have been developed by Rao (1967). In the same paper,

he discusses the possibility of using a subset of covariates for covariance adjusted estimator to obtain more efficient estimators of $\underset{\sim}{\beta}$ and setting confidence bands for some linear function of $\underset{\sim}{\beta}$. He obtains the necessary and sufficient condition for the unweighted least square estimator to be the best linear unbiased for $\theta$ in a generalised growth curve model. The condition is

$$\Omega = X \Gamma X' + Z \Delta Z' + \sigma^2 I , \qquad (1.2.1)$$

where $Z$ is any $p \times (p-q)$ matrix of rank $p-q$ such that $X'Z = 0$ and $\Gamma$ and $\Delta$ are arbitrary unknown positive definite matrices. Grizzle and Allen (1969) note that the Potthoff-Roy's estimator with $G = S^*$, where

$$NS^* = Y \left[ I - W'(WW')^{-1}W \right] Y' ,$$

Rao's (1966) estimator with all covariates, and Khatri's (1966) estimator are identical. Grizzle and Allen (1969) also consider the issue of selection of covariates.

Swamy (1971) considers a more general version of RCR model by allowing X to be different for each unit but the dimension of X is the same for all the units. He derives a two-step estimation procedure. First, the individual regression parameters are estimated by the least square method. Then the estimates of the population regression parameters and variances are obtained. This procedure may sometimes give a negative estimate of variance. The estimators are consistent as the length of the series becomes large.

A model of considerable interest for our work is the one where he assumes that the errors follow a stationary autoregressive process of order 1 (Swamy, 1971, Section 4.4). He considers that each unit has its own autocorrelation coefficient and obtains consistent estimators as the length of the series becomes large.

In a recent paper, Carter and Yang (1986) consider a RCR model where the length of each series of observations may not be the same. They obtain consistent estimator of $\underset{\sim}{\beta}$, which is modification of estimator given by Swamy (1971), for this case when the smallest of the lengths of series may be small but number of units is large.

Khatri (1973) derives the likelihood ratio test for testing three types of dispersion structures viz., (i) independence of two sets, (ii) the sphericity of the dispersion matrix and (iii) intraclass model of the dispersion matrix. He also obtains the null distribution of the test statistics.

Bayesian approach to growth curves is considered by Geisser (1970), Lee and Geisser (1972, 1975) and Fearn (1975, 1977). Geisser (1970) provides Bayesian justification for the covariance adjusted estimator of Rao (1967). Lee and Geisser (1972) recommend the use of posterior mode rather than posterior mean as it considerably simplifies the computation procedure. The likelihood ratio test for

$H_o : \Omega = X \Gamma X' + Z \Delta Z' + \sigma^2 I$ ($\Gamma$ and $\Delta$ positive definite and unknown)

vs

$H_1 : \Omega$ arbitrary

has also been developed. Fearn (1975) views the RCR model as a hierarchical linear model and performs a Bayesian analysis. A case when population growth curve is of a lower dimension than the individual growth curves is dealt with in a later paper by Fearn (1977).

Rao (1975) has considered empirical Bayes (EB) estimators for RCR model for estimation of coefficients $\beta_i$'s when $F$, $\beta$ and $\sigma^2$ are unknown. It may be noted that the estimators are not EB in the strict sense of the word because the estimators of $\beta_i$'s are obtained by replacing the parameters by their sample estimates. In the rigid sense (e.g. Maritz, 1970), the parameters have to be estimated from the past occurrences of the same situation. Rao (1975) makes a very thoughtful suggestion of introducing a shrinkage parameter, to be determined suitably, when the unknown parameters are replaced by their estimates. The value of such a parameter which minimizes the average mean square error (MSE), where average is taken over all $\beta_i$'s in the sample, has been obtained. Strenio et al. (1983) consider a Potthoff-Roy type of model for mean and structured dispersion matrix of RCR model. In their model, individual regression parameters are assumed to be functions of some fixed known covariates which may be different for different individuals.

In a series of papers, Reinsel (1982, 1984a, 1985) has considered generalisation of growth curve for one characteristic to a multicharacteristic case. Reinsel (1982) considers generalisation of Potthoff-Roy model to multicharacteristic case and develops procedures for testing the hypothesis about various linear combinations of elements of fixed effect parameters. A similar generalisation of RCR model has been considered by Reinsel (1984a). The choice of optimum value of constant multiplier when estimated values are used in place of unknown parameters and approximate prediction MSE have also been considered. A case when design matrix may not be the same for all the individuals is considered by Reinsel (1985). An expression for approximate prediction MSE has also been derived.

Reinsel (1982) shows that when the dispersion matrix is of the form

$$\Omega = (X \otimes I) \Gamma (X' \otimes I) + (Z \otimes I) \Delta (Z' \otimes I) + (I \otimes V_e),$$

$$(1.2.2)$$

where $\otimes$ denotes a Kronecker product, which is a multivariate generalisation of (1.2.1), the unweighted least square estimators are the best linear unbiased. A likelihood ratio test for testing the structure (1.2.2) of dispersion matrix is derived by Chinchilli and Carter (1984).

Anderson (1978) considers a multivariate autoregressive process for repeated series and obtains the ML estimates of the

parameters. Procedures to test the nature of autoregressive parameter matrix are also developed.

Laird and Ware (1982) consider a family of two-stage models for longitudinal studies. Use of EM algorithm for ML and restricted ML estimation procedure is explained. This is illustrated by examples.

A priori reduction in the number of covariates for models with different structures of the dispersion matrix is considered by Verbyla (1986). The reduction in number of covariates is shown to depend on the structure in inverse of dispersion matrix and the form of the design matrix. In particular, the cases of autoregressive processes and RCR models are considered.

For more details on models used in longitudinal studies, one may refer to Geisser (1980), Woolson and Leeper (1980), Khatri (1985) and Ware (1985).

1.2.2 <u>Prediction</u> : Before we move on to review the work on prediction in longitudinal studies, it would be proper to make the distinction between estimation and prediction. Estimation is a procedure of determining the value of fixed parameters whereas prediction is concerned regarding the value about the random event itself. Prediction in time series literature is commonly known as forecasting. A good review on some early procedures of forecasting in univariate time series is given by

Newbold and Granger (1974). However, in the present thesis we are not interested in time series techniques.

Some work on prediction in linear models is also reported in literature. Consider the model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{e} \ .$$

where $\underset{\sim}{y}$ is a $px1$ vector of observable quantities, $\underset{\sim}{\beta}$ is a $qx1$ vector of regression parameters and $\underset{\sim}{e}$ is a $px1$ vector of errors. Assume that

$$var(\underset{\sim}{y}) = \sigma^2 V_{11} \ .$$

The purpose is to predict $y_*$ as a linear function of $\underset{\sim}{y}$, where

$$Y_* = \underset{\sim *}{x}' \ \underset{\sim}{\beta} + e_* \ ,$$

and $\underset{\sim *}{x}$ is a known $qx1$ vector, $var(y_*) = \sigma^2 V_{22}$ and $\quad$ ,
$cov(\underset{\sim}{y}, y_*) = \sigma^2 \underset{\sim}{V}_{12}$.

Let $\underset{\sim}{\ell}' \ \underset{\sim}{y}$ be a linear predictor of $y_*$ i.e.

$$\hat{Y}_* = \underset{\sim}{\ell}' \ \underset{\sim}{y} = \underset{\sim}{\ell}' \ X \ \underset{\sim}{\beta} + \underset{\sim}{\ell}' \ \underset{\sim}{e}.$$

It can be seen that $\hat{y}_*$ is unbiased for $\underset{\sim *}{x}'\underset{\sim}{\beta}$, the expected value of $y_*$, if

$$X'\underset{\sim}{\ell} = \underset{\sim *}{x} \ . \qquad\qquad (1.2.3)$$

The variance of $\hat{y}_*$ is given by

$$var(\hat{y}_*) = E(y_* - \underset{\sim *}{x}' \ \underset{\sim}{\beta})^2$$
$$= \sigma^2 \ \underset{\sim}{\ell}' \ V_{11} \underset{\sim}{\ell} \ .$$

This measures the mean square error (MSE) of $y$ from its mean $\underset{\sim}{\ell}' X \underset{\sim}{\beta}$. However, a more relevant measure of deviation of $y_*$ from its predictor $\hat{y}_*$ is given by

$$E(\hat{y}_* - y_*)^2 = E[(\hat{y}_* - \underset{\sim}{\ell}' X \underset{\sim}{\beta}) + (\underset{\sim}{\ell}' X \underset{\sim}{\beta} - y_*)]^2$$

$$= \sigma^2 [\underset{\sim}{\ell}' V_{11} \underset{\sim}{\ell} + V_{22} - 2\underset{\sim}{\ell}' \underset{\sim}{V}_{12}]. \quad (1.2.4)$$

Hence the best linear unbiased predictor is obtained when $\underset{\sim}{\ell}$ is chosen in such a way that it minimizes (1.2.4) subject to (1.2.3). It is easily seen that the best choice of $\underset{\sim}{\ell}$ is given by

$$\underset{\sim}{\ell} = V_{11}^{-1} \underset{\sim}{V}_{12}.$$

The predictor is thus obtained as

$$\hat{y}_* = \underset{\sim}{x}'_* \underset{\sim}{\beta} + \underset{\sim}{V}'_{12} V_{11}^{-1}(\underset{\sim}{y} - X \underset{\sim}{\beta})$$

$$= E(y_* | \underset{\sim}{y}), \quad (1.2.5)$$

The predictor $\hat{y}_*$ is called the best linear unbiased predictor of $y_*$. If $\underset{\sim}{\beta}$ and $V$ are not known then it is natural to put the estimated values of these parameters in (1.2.5). There will be an increase in MSE due to the use of estimated values in place of the parameters. This type of predictor is considered by Goldberger (1962).

To get some general idea of the work done on prediction in linear models one may refer to Bibby and Toutenburg (1977).

A problem of considerable interest in longitudinal studies is of prediction. The work done on multivariate missing

data problems and problem of growth curves are the two specific instances which show the importance of prediction in such studies. Some early work is due to Lee and Geisser (1972,1975), Rao (1975) and Fearn (1975). Since then it has received considerable attention of the researchers, for example see Rao (1977,1981), Geisser (1981), Copas (1983), Reinsel (1984a,b,c; 1985), Rao and Boudreau (1985) and references therein.

Suppose that N units are observed at p time points and observations are represented by p component vectors $\underset{\sim}{y}_i$ (i = 1,2,...,N). Two types of prediction problems are considered in literature. For (N+1)th unit, vector $\underset{\sim}{y}_{N+1}$ is observed at $p_1 (< p)$ time points and we wish to predict the remaining $p_2 (= p-p_1)$ values. This is called the conditional prediction. Another type of problem is of predicting the values of measurement at (p+h)th time point for these N units. This is called the problem of predicting the future observations.

The problem of conditional prediction for the generalised growth curve model of Potthoff and Roy (1964) is treated analytically by Lee and Geisser (1972). Computational feasibility of predictors based on various approaches is discussed in detail by Lee and Geisser (1975). This is illustrated numerically by working out two examples.

Rao (1975) and Fearn (1975) also consider the problem of conditional prediction in RCR models when the parameters are known. Rao (1975) suggests the use of estimates of the parameters

when the parameters are unknown and obtains MSE of prediction when they are known. This will give an underestimate of actual MSE because it does not take into account the variability due to the use of estimates of the parameters.

Reinsel (1984a) considers prediction in a multivariate random effects model when the random coefficients depend on some covariates. An expression for approximate prediction MSE, in caseswhere estimates of unknown parameters are used, is derived. Reinsel (1984b) considers this problem in a multivariate linear model with general dispersion matrix and obtains the prediction MSE of the usual least square predictor after using estimates in place of the parameters. He shows that the increase in prediction MSE due to use of estimates is substantial if the sample size is small. This work is extended by Reinsel (1984c) where he derives an approximate prediction MSE in the generalised multivariate linear model when $\varrho$ has a linear structure. Models considered by Rao (1965), Khatri (1966) and Grizzle and Allen (1969) are particular cases of this model.

Rao (1977, 1981) considers the problem of prediction for future time points. Rao (1977) considers prediction in three situations viz., 1) prediction for a particular unit, 2) prediction for a unit drawn from a specified population and 3) simultaneous prediction for a number of units, in a linear model. The performance of various methods, such as the best linear unbiased predictors, empirical best linear predictor and the best homogeneous linear predictors, are compared by applying

them to the data of Elston and Grizzle (1962). Rao (1981) considers the simultaneous prediction of future observations on a number of units in a polynomial growth curve model using their past values.

Rao and Boudreau (1985) consider the problem of predicting future observations in factor analytic model. In such models the prior knowledge of orthogonal functions is not assumed and both the orthogonal functions and factor variables have to be estimated. They also discuss a method of selecting an appropriate model.

A model free data analytic approach to conditional prediction is considered by Geisser (1981). A detailed study of issues involved in prediction and shrinkage in the regression models is considered by Copas (1983). Subset selection in regression analysis is also considered.

A potentially useful method of prediction is based on Kalman filters. Kalman filter is a recursive, unbiased least squares estimator of a Gaussian random signal (Wegman, 1983). Since most of the work on Kalman filter, including the original work of Kalman (1960), is reported in engineering literature it is not very popular among statisticians. Harrison and Stevens (1971, 1976) use Kalman filter for short term forecasting. Duncan and Horn (1972) observe similarity between RCR model and Kalman filter model. The recursive estimates occurring in the Kalman filter models are seen in the terms of regression theory.

Meinhold and Singpurwalla (1983) demonstrate that Kalman
filters are very much similar to linear models. Connection
between Kalman filters and Goldberger-Theil estimator is shown
by Diderrich (1985). Sallas and Harville (1981) use the Kalman
filter technique for recursive estimation in mixed linear models.

1.3 MOTIVATION

From the review of available literature it is clear
that most of the researchers have assumed a prior knowledge
about the degree of polynomial or the dimension of the regression
parameters. Among some of the exceptions are the works of
Rao (1981) and Lee and Tan (1984) for fixed parameter's case.
Rao's (1981) objective function for selection of the degree
of the polynomial is the minimisation of the compound mean
square prediction error (CMSPE). Lee and Tan (1984) consider
this problem in a Potthoff-Roy model by a Bayesian method and
a method based on likelihood ratio test. Though RCR models
are being used extensively, it seems, no work has been done on
the selection of the degree of the polynomial (or the dimension
of the dispersion matrix of random regression coefficients).
This is considered in Chapter II of the thesis. Prediction
under the uncertainty about the degree of the polynomial in RCR
model is also considered. The importance of covariate selection
for covariance adjusted estimators is well recognised. The
most widely used method of covariate selection is the one
proposed by Rao (1967). Grizzle and Allen (1969) have used the

method in the analysis of some data. This method is not based
on any optimality criterion. Two strategies which are based on
A- and D-optimality criteria used in design of experiments are
proposed in this chapter.

Strenio (1981) and Strenio et al. (1983) consider a
growth curve model where they assume that the regression parameters
are functions of fixed known covariates. The mean has structure
of Potthoff-Roy model and dispersion matrix has the structure
of RCR model. This case has been generalised to multicharacteris-
tic case by Reinsel (1982,1984a) which is a very significant
contribution in the analysis of longitudinal data. For simplicity
the polynomials of the same degree have been fitted for all
the characteristics, and the set of covariates is common for
all the characteristics. This may not be the most parsimonious
way of fitting the model. Moreover, there is a need to relax
these conditions as will be illustrated by an example in
preharvest forecasting. These considerations have motivated
us for the work done in Chapter III.

Regression models using environmental factors as
independent variables are widely used for forecasting crop yield.
The regression coefficient of such a fit, in a way, represent
the 'state' of the environment which is not directly observable.
It can be measured by observing the yield. Now, due to industrial
activity and other reasons such as fumes and pollutants being
thrown in the atmosphere, there may be deterioration in the

'state' of environment over the years. This change may be represented by an autoregressive process similar to that of Anderson (1978). Thus there is a need to combine these two approaches simultaneously to form a two-stage model. In the first stage a usual regression model is considered and in the second stage the regression parameters are considered to have an autoregressive structure. Such a formulation is similar to Kalman filter type of models and is the topic of Chapter IV.

In most of the work on RCR models it is assumed that the errors are independently distributed or distributed as $N(\underset{\sim}{0}, \sigma^2 V)$ with a known $V$, for example see Rao (1975). If $V$ is a general unknown matrix then it may not be possible to estimate $V$ because all the parameters may not be identifiable. If $V$ has a suitable structure so that the parameters are identifiable then it may be possible to estimate them. In some cases it may be proper to assume that the errors of the same unit show an autoregressive pattern. Swamy (1971, Section 4.4) and Mansour et al. (1985) consider such structures. Swamy (1971) considers a stationary autoregressive process of order 1 but allows the autocorrelation coefficient to change from unit to unit. His analysis is valid if the length of the individual series is large which may not be a realistic assumption in a longitudinal study. Mansour et al. (1985) consider a nonstationary autoregressive process but the dispersion matrix of individual regression parameters is not very general. In Chapter V a mixed linear

model with the errors following a stationary autoregressive process of order 1 with the same autocorrelation coefficient for all the units is considered.

## 1.4   EM ALGORITHM

The concept EM algorithm is due to Dempster et al.(1977). It is an iterative technique of obtaining the ML estimates. The algorithm is so named because it consists of two steps, the expectation(E) step and the maximization (M) step. As such, an algorithm is a special process of solving certain type of problems.   The EM algorithm does not actually specify the sequence of steps to be performed and so in the strict sense of the word it is not an algorithm but a procedure or a technique.

Some of the reasons for popularity of this technique among the statisticians are the following :

1.      It achieves synthesis for ML estimation, which is a very satisfactory estimation procedure, by exhibiting various methods as special cases of the general EM algorithm. Wide range of problems fall within periphery of this algorithm.   Thus this is a remarkable procedure.

2.      The general EM theory, which shows that each iteration increases the likelihood, can be applied.

3.      The expressions for defining iterative steps of the algorithm have meaningful statistical interpretations. Thus it is appealing.

In the EM setting the observed data are broadly seen as incomplete data coming from a larger system of quantities. When we talk of complete and incomplete data we visualise two sample spaces (say $\mathcal{X}$ and $\mathcal{Y}$) and a many-one function from $\mathcal{X}$ to $\mathcal{Y}$, i.e. for an observed $\underset{\sim}{y} \in \mathcal{Y}$, $\underset{\sim}{x}$ is not known completely. What is known is that $\underset{\sim}{x}$ will lie in some region determined by the observed $\underset{\sim}{y}$ (say $\mathcal{X}(\underset{\sim}{y})$). Putting differently, $\underset{\sim}{x}$ is not observed directly but only through $\underset{\sim}{y}$. We say that $\underset{\sim}{x}$ is complete data and $\underset{\sim}{y}$ is incomplete data.

The family of complete data, say $f(\underset{\sim}{x}|\underset{\sim}{\varphi})$, is assumed to be known. The incomplete data specification family $g(\underset{\sim}{y}|\underset{\sim}{\varphi})$ is derived so that

$$g(\underset{\sim}{y}|\underset{\sim}{\varphi}) = \int_{\mathcal{X}(\underset{\sim}{y})} f(\underset{\sim}{x}|\underset{\sim}{\varphi}) \, d\underset{\sim}{x}. \qquad (1.4.1)$$

EM algorithm aims at locating the value of $\underset{\sim}{\varphi}$ which maximizes $g(\underset{\sim}{y}|\underset{\sim}{\varphi})$ given the observed $\underset{\sim}{y}$. This maximization is achieved by iteration and each iteration consists of two steps; the E-step and the M-step.

Dempster, Laird and Rubin (1977) have shown that the EM algorithm can be applied to any family of distribution functions. But for the sake of simplicity in understanding the attention is restricted to the exponential family. Let $f(\underset{\sim}{x}|\underset{\sim}{\varphi})$ belong to a regular exponential family of distributions i.e.

$$f(\underset{\sim}{x}|\underset{\sim}{\varphi}) = b(\underset{\sim}{x}) \cdot \exp\left[\underset{\sim}{\varphi}'(\underset{\sim}{t}(\underset{\sim}{x}))\right]/a(\underset{\sim}{\varphi}).$$

It is well-known that $t(x)$ is a complete sufficient statistic. The problem of EM algorithm may be stated as follows. We are given the family of complete data, an observed $y$ and a many-one function from $X$ to $Y$. The aim is to obtain the value of that maximizes $g(y|\varphi)$ subject to the condition (1.4.1).

Let $\varphi^{(p)}$ be the current value of $\varphi$ after p-iterations. The E-step estimates the complete sufficient statistic $t(x)$ by finding the posterior expected value of the sufficient statistic, i.e.,

$$t^{(p)} = E\left[t(x)|y, \varphi^{(p)}\right].$$

The M-step simulates ML estimates which would have been arrived at if the complete data were known. But the data are incomplete and M-step can be carried out only in a proxy. The proxy is produced by E-step. The M-step obtains $\varphi^{(p+1)}$ as a solution of

$$E(t(x)|\varphi) = t^{(p)}.$$

This is the usual form of normal equations for obtaining ML estimates given the data from an exponential family of distributions.

In general EM algorithm, no reference is made to exponential family. Dempster et al., (1977) have proved that a general EM algorithm produces a sequence that results in non-decreasing value likelihood function and that a convergence is achieved.

Like any other method of numerical maximization the EM method can also converge to local maxima rather than the global maxima. The conditions for existence of unique maxima are well known. In the presence of local maxima the choice of initial value is very crucial. In cases considered in this thesis the starting values are obtained by simple estimators which have some desirable properties such as consistency and in some cases unbiasedness. Moreover, random checks are also given by other numerical maximizing algorithms and in all cases considered they converged to the same values as given by EM algorithm. The time taken by EM to converge is much smaller in comparison to other methods. Moreover, EM has certain other advantages such as meaningful statistical interpretation of each step and increasing nature of the likelihood at each iteration.

Though the EM algorithm is a powerful concept for simplifying the computations of ML estimates of the parameters there are some limitations. It locates the region of convergence very quickly but after that the rate of convergence is slow. It is not easy to get the variances of the estimates so obtained.

Dempster et al. (1981) consider estimation via EM algorithm in linear variance components models. Strenio (1981) uses this method for estimation of variances in hierarchical linear models. Laird and Ware (1982) use EM algorithm for estimation in mixed linear models.

# CHAPTER II

## PREDICTION UNDER UNCERTAINTY OF DEGREE OF POLYNOMIAL IN RANDOM COEFFICIENT REGRESSION MODELS

### 2.1 INTRODUCTION

While fitting a polynomial to the data it is usually assumed that the degree of the polynomial is known. Misspecification in the case of fixed coefficient regression models has been considered in the literature by Hocking (1974) and references therein. The topic is also related to selection of variables for prediction on which a vast literature is available (Allen, 1971 and Hocking, 1976). Some work on determining the degree of polynomial in growth curve models has also been reported in literature. For fixed parameter case, Rao (1981) studies this problem from the point of view of minimizing the compound mean square prediction error (CMSPE). Lee and Tan (1984) consider the problem of estimation of degree of polynomial in a general growth curve model by a Bayesian method and a method based on the likelihood ratio test. Some work on the factor analysis type of growth curve model has been reported by Rao and Boudreau (1985). In this chapter, we consider the optimum choice of degree of polynomial and the problem of prediction under the uncertainty about the degree of the polynomial in random coefficient regression (RCR) models. Some strategies for the optimum choice of covariates are also

discussed. Some results of this chapter have been published in Oza and Shukla (1986).

The format of this chapter is as follows. In Section 2.2 we propose a RCR model and consider two situations which may arise due to uncertainty about the model. Section 2.3 gives the mean square error (MSE) of predictors for the two situations considered in Section 2.2 and Section 2.4 contains comparison of these errors. In Section 2.5 we consider the effect of using estimated parameters for prediction and compare the performance of optimum predictors with other similar predictors by a cross validation study on a real set of data. The problem of selection of degree of polynomial is considered in Section 2.6. Section 2.7 contains some results on optimum choice of covariates. General discussion and summary are given in Section 2.8.

## 2.2 MODEL

Let us consider an observable p-vector random variable $\underset{\sim}{y}_i$, $i = 1, 2, \ldots, N$, which can be expressed as

$$\underset{\sim}{y}_i = X \underset{\sim}{\beta}_i + \underset{\sim}{e}_i , \quad i = 1, 2, \ldots, N; \qquad (2.2.1)$$

where X is a known design matrix of order $p \times q$, which is same for all i, $\underset{\sim}{\beta}_i$ is a $q \times 1$ vector of random regression parameters and $\underset{\sim}{e}_i$ is an independent $p \times 1$ error vector with mean zero and dispersion matrix $\sigma^2 I$.

For random vector $\underset{\sim}{\beta}_i$ we assume that

$$\underset{\sim}{\beta}_i = \underset{\sim}{\beta}^* + \underset{\sim}{\eta}_i \ , \ i = 1, 2, \ldots, N;$$

with $E(\underset{\sim}{\beta}_i) = \begin{array}{c} q_1 \\ q-q_1 \end{array} \left[ \begin{array}{c} \underset{\sim}{\beta}_1 \\ \underset{\sim}{0} \end{array} \right] = \underset{\sim}{\beta} \ ; \ var(\underset{\sim}{\beta}_i) = var(\underset{\sim}{\eta}_i) = \begin{array}{c} q_1 \\ q-q_1 \end{array} \left[ \begin{array}{cc} F_{11} & F_{12} \\ F_{21} & F_{22} \end{array} \right].$

$$\begin{array}{cc} q_1 & q-q_1 \end{array}$$

$$(2.2.2)$$

Partition $X$ and $\underset{\sim}{\beta}_i$ as $X = (X_1 : X_2)$, where $X_1$ and $X_2$ are $p \times q_1$ and $p \times (q-q_1)$ matrices, respectively, and $\underset{\sim}{\beta}_i = (\beta'_{\sim 1i} : \beta'_{\sim 2i})'$ where $\underset{\sim}{\beta}_{1i}$ and $\underset{\sim}{\beta}_{2i}$ are $q_1 \times 1$ and $(q-q_1) \times 1$ vectors, respectively. Then the model (2.2.1) can be written as

$$\underset{\sim}{y}_i = X_1 \underset{\sim}{\beta}_{1i} + X_2 \underset{\sim}{\beta}_{2i} + \underset{\sim}{e}_i$$

$$= X_1 \underset{\sim}{\beta}_1 + X \underset{\sim}{\eta}_i + \underset{\sim}{e}_i \ , \ i = 1, 2, \ldots, N.$$

In this model the dimension of population regression parameter $\underset{\sim}{\beta}_1$ is smaller than for individual regression parameter $\underset{\sim}{\beta}_i$'s. Such a model has been considered by Fearn (1977).

In the present study we shall consider two situations which may arise due to uncertainty about the model.

Case 1 : Here one considers an incorrect dispersion matrix of $\underset{\sim}{\beta}_i$ by assuming (2.2.2) as

$$E_1(\underset{\sim}{\beta}_i) = \left[ \begin{array}{c} \underset{\sim}{\beta}_1 \\ \underset{\sim}{0} \end{array} \right] \ , \ var_1(\underset{\sim}{\beta}_i) = \left[ \begin{array}{cc} F_{11} & 0 \\ 0 & 0 \end{array} \right] \qquad (2.2.3)$$

In other words, in this case one wrongly considers the dimension of individual regression parameters same as that of population regression parameters, i.e. $q_1$.

Case 2: Here one considers an incorrect dimension of the population regression parameters by assuming (2.2.2) as

$$E_2(\underset{\sim}{\beta}_i) = \underset{\sim}{\beta} = \begin{bmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{bmatrix} \; ; \; \text{var}_2(\underset{\sim}{\beta}_i) = F. \qquad (2.2.4)$$

In this case one wrongly considers the degree of population parameters to be same as that for individual regression parameters, i.e. q.

In the above cases $E_1$ and $\text{var}_1$ , $E_2$ and $\text{var}_2$ denote expected values and variances corresponding to cases 1 and 2, respectively, while under the correct model they are denoted by E and var as given in (2.2.2).

## 2.3 PREDICTORS AND THEIR ERRORS

2.3.1 Case 1 : Let $\underset{\sim}{\tilde{\beta}}_{1i}$ be the least square estimator of $\underset{\sim}{\beta}_1$ from the i-th individual i.e. $\underset{\sim}{\tilde{\beta}}_{1i} = (X_1' X_1)^{-1} X_1' \underset{\sim}{y}_i$. If $\sigma^2, \underset{\sim}{\beta}_1$ and $F_{11}$ are known, then the optimum linear predictors of $\underset{\sim}{t}' \underset{\sim}{\beta}_{1i}$ is $\underset{\sim}{t}' \underset{\sim}{\beta}_{1i}^b$ (Rao, 1975) where

$$\underset{\sim}{\beta}_{1i}^b = \underset{\sim}{\tilde{\beta}}_{1i} - \sigma^2 U_1 (F_{11} + \sigma^2 U_1)^{-1} (\underset{\sim}{\tilde{\beta}}_{1i} - \underset{\sim}{\beta}_1), \qquad (2.3.1)$$

and $U_1 = (X_1' X_1)^{-1}$.

The above estimator $\underset{\sim}{\beta}_{1i}^b$ is also a Bayes estimator for an appropriate prior. The minimum mean square error (MMSE) predictor for a future observation on (N+1)-th individual, given first p observations $\underset{\sim}{y}_{N+1}$, corresponding to a point $\underset{\sim}{z}'_{01}$

(belonging to row space of $X_1$), is given by

$$\tilde{Y}_{N+1} = z'_{01}[\tilde{\beta}_{1N+1} - \sigma^2 U_1(F_{11}+\sigma^2 U_1)^{-1}(\tilde{\beta}_{1N+1}-\beta_1)].$$

$$(2.3.2)$$

Since the values of parameters $\sigma^2$, $\beta_1$ and $F_{11}$ are not known we substitute their estimated values based on $Y_1, Y_2, \cdots, Y_N$. One might introduce a constant multiplier $c$ in the second term of $(2.3.2)$ when parameter values are replaced by their estimated values, as done by Rao (1975), and choose the value of $c$ which minimizes MSE. In the present case, such an optimum $c$ depends on unknown F, unlike that of Rao (1975), and has not been introduced here. However, introduction of such a constant $c$ has been considered for model of case 2 in section 2.5 where it does not depend upon unknown parameters. There is some loss of information due to not using $Y_{N+1}$ for estimation of parameters. The expressions for prediction errors are difficult to obtain when information on (N+1)-th individual is also used. It is expected that the loss due to not including $Y_{N+1}$ in estimation is not serious when N is moderately large.

We shall assume that an unbiased estimator of $\sigma^2$, denoted by $\hat{\sigma}^2$, is available. This may be obtained by fitting a higher degree polynomial to individual's data and considering pooled residual sum of squares $w_1$ distributed as $\sigma^2 \chi^2(f)$.

When $\beta_1$ and $F_{11}$ are unknown we may estimate them by,

$$N\bar{\underset{\sim}{\beta}}_1 = \sum_{i=1}^{N} \tilde{\underset{\sim}{\beta}}_{1i} \; ,$$

$$(N-1)\,\widetilde{(F_{11}+\sigma^2 U_1)} = \sum_{i=1}^{N} (\tilde{\underset{\sim}{\beta}}_{1i}-\bar{\underset{\sim}{\beta}}_1)(\tilde{\underset{\sim}{\beta}}_{1i}-\bar{\underset{\sim}{\beta}}_1)' = B_1.$$

These estimators are the maximum likelihood estimators of $\underset{\sim}{\beta}_1$ and $F_{11}+\sigma^2 U_1$, respectively, for the Case 1. Substituting these estimators in $(2.3.2)$ we get a predictor of $Y_{N+1}$ as

$$\tilde{Y}_{N+1*} = \underset{\sim}{z}'_{1}\left[\tilde{\underset{\sim}{\beta}}_{1N+1}-(N-1)\,\hat{\sigma}^2 U_1 B_1^{-1}(\tilde{\underset{\sim}{\beta}}_{1N+1}-\bar{\underset{\sim}{\beta}}_1)\right]. \qquad (2.3.3)$$

If $\underset{\sim}{\eta}_i$ is distributed as a multivariate normal distribution then $B_1$ is distributed as $W(P;N-1,q_1)$, where

$$P = U_1 X_1'(XFX' + \sigma^2 I)X_1 U_1.$$

We assume that $w_1$ and $B_1$ are independently distributed. We also note that if $B_1 \sim W(P;N-1,q_1)$ then,

$$E(B_1^{-1}PB_1^{-1}) = (N-2)\{(N-q_1-1)(N-q_1-2)(N-q_1-4)\}^{-1}P^{-1}.$$

Using the above results and noting that under the true model

$$Y_{N+1} = \underset{\sim}{z}'_{1}\,\underset{\sim}{\beta}_1 + \underset{\sim}{z}'_{0}\,\underset{\sim}{\eta}_{N+1} + \underset{\sim}{e}_{N+1} \; ,$$

where $\underset{\sim}{z}'_{0} = (\underset{\sim}{z}'_{01}:\underset{\sim}{z}'_{02})$ belongs to row space of $X$, we obtain

$$MSE(\tilde{Y}_{N+1*}) = E(Y_{N+1} - \tilde{Y}_{N+1*})^2$$

$$= \sigma^2 + \sigma^2 \underset{\sim}{z}'_{01} U_1 \underset{\sim}{z}_{01} + N_1(q_1)\sigma^4 \underset{\sim}{z}'_{01} U_1 P^{-1} U_1 \underset{\sim}{z}_{01} + \underset{\sim}{v}' F_{22}\underset{\sim}{v}$$

$$+ 2N_2(q_1)\,\sigma^2 \underset{\sim}{z}'_{01} U_1 P^{-1}(F_{12}+U_1 X_1' X_2 F_{22})\underset{\sim}{v} \qquad (2.3.4)$$

where, for any integer $k$,

$$N_1(k) = N_2(k) \left[ (f+2)(N^2-1)(N-2) \{fN(N-k-1)(N-k-4)\}^{-1} - 2 \right],$$

$$N_2(k) = (N-1)(N-k-2)^{-1}, \quad \underset{\sim}{v} = (\underset{\sim}{z}_{o2} - X_2' X_1 U_1 \underset{\sim}{z}_{o1}).$$

2.3.2 <u>Case 2</u> :  Let $\underset{\sim}{\beta}_i$ be the least square estimator of $\underset{\sim}{\beta}$ from the i-th individual i.e. $\underset{\sim}{\hat{\beta}}_i = (X'X)^{-1} X' \underset{\sim}{y}_i$.  The MMSE predictor for a future observation on (N+1)-th individual, given first p observations $\underset{\sim}{y}_{N+1}$, corresponding to the point $\underset{\sim}{z}_o$, is given by

$$\hat{y}_{N+1} = \underset{\sim}{z}_o' \left[ \underset{\sim}{\hat{\beta}}_{N+1} - \sigma^2 U (F+\sigma^2 U)^{-1} (\underset{\sim}{\hat{\beta}}_{N+1} - \underset{\sim}{\beta}) \right], \qquad (2.3.5)$$

where $U = (X'X)^{-1}$.

When $\underset{\sim}{\beta}$ and $F$ are unknown we estimate them by $N\underset{\sim}{\bar{\beta}} = \Sigma \underset{\sim}{\hat{\beta}}_i$, where the summation extends over all N individuals, and

$$(N-1)(\hat{F+\sigma^2 U}) = \sum_{i=1}^{N} (\underset{\sim}{\hat{\beta}}_i - \underset{\sim}{\bar{\beta}})(\underset{\sim}{\hat{\beta}}_i - \underset{\sim}{\bar{\beta}})' = B.$$

Again these estimators are the maximum likelihood estimators for the respective parameters for Case 2.

Substituting these estimators in (2.3.5) we get a predictor of $y_{N+1}$ as

$$\hat{y}_{N+1*} = \underset{\sim}{z}_o' \left[ \underset{\sim}{\hat{\beta}}_{N+1} - (N-1) \hat{\sigma}^2 U B^{-1} (\underset{\sim}{\hat{\beta}}_{N+1} - \underset{\sim}{\bar{\beta}}) \right].$$

Under the assumption of multivariate normality
$B \sim W(F + \sigma^2 U; N-1, q)$ and thus we obtain

$$MSE(\hat{y}_{N+1*}) = E(y_{N+1} - \hat{y}_{N+1*})^2$$

$$= \sigma^2 + \sigma^2 \underset{\sim}{z}_o' U \underset{\sim}{z}_o + N_1(q)\sigma^4 \underset{\sim}{z}_o' U (F+\sigma^2 U)^{-1} U \underset{\sim}{z}_o. \qquad (2.3.6)$$

## 2.4 COMPARISON OF PREDICTION ERRORS

We are interested in comparing the efficiencies of predictors $\tilde{y}_{N+1*}$ and $\hat{y}_{N+1*}$. Upto this time we have derived expressions for MSE's for any N and for a general regression setup. However, these expressions are complicated and to get some insight we shall consider the case of orthogonal polynomial when N is large. Under this situation the expressions (2.3.4) and (2.3.6) simplify to

$$MSE(\tilde{y}_{N+1*}) = \sigma^2 + \sigma^2 \underset{\sim}{z}'_{o1} \underset{\sim}{z}_{o1} - \sigma^4 \underset{\sim}{z}'_{o1}(F_{11}+\sigma^2 I)^{-1} \underset{\sim}{z}_{o1} + \underset{\sim}{z}'_{o2} F_{22} \underset{\sim}{z}_{o2}$$

$$+ 2\sigma^2 \underset{\sim}{z}'_{o1}(F_{11}+\sigma^2 I)^{-1} F_{12} \underset{\sim}{z}_{o2}.$$

$$MSE(\hat{y}_{N+1*}) = \sigma^2 + \sigma^2 \underset{\sim}{z}'_{o} \underset{\sim}{z}_{o} - \sigma^4 \underset{\sim}{z}'_{o}(F+\sigma^2 I)^{-1} \underset{\sim}{z}_{o}.$$

It is not difficult to see after some manipulation that

$$MSE(\tilde{y}_{N+1*}) - MSE(\hat{y}_{N+1*})$$

$$= \sigma^2 \underset{\sim}{z}' \left[ \left( \sigma^2 \begin{bmatrix} \Gamma'_{2.1} \\ I \end{bmatrix} V^{-1}_{2.1} \left[ \Gamma_{2.1} : I \right] + \begin{bmatrix} 0 & \Gamma'_{2.1} \\ \Gamma_{2.1} & \sigma^{-2} F_{22} - I \end{bmatrix} \right] \underset{\sim}{z}_{o},$$

$$(2.4.1)$$

where $\Gamma_{2.1} = F_{21}(F_{11}+\sigma^2 I)^{-1}$ is the matrix of regression coefficients of $\hat{\underset{\sim}{\beta}}_{2i}$ on $\hat{\underset{\sim}{\beta}}_{1i}$ where $\hat{\underset{\sim}{\beta}}_i = (\hat{\underset{\sim}{\beta}}'_{1i} : \hat{\underset{\sim}{\beta}}'_{2i})'$ and $V_{2.1}$ is the conditional dispersion of $\hat{\underset{\sim}{\beta}}_{2i}$ given $\hat{\underset{\sim}{\beta}}_{1i}$, given by

$$V_{2.1} = F_{22} + \sigma^2 I - F_{21}(F_{11}+\sigma^2 I)^{-1} F_{12}.$$

The first matrix of (2.4.1) is always positive definite and the second matrix is indefinite and nothing can be said in general. However, when $\Gamma_{2.1}$ is small and $F_{22}$ has large characteristic roots then the second matrix is likely to be positive semidefinite and $\hat{y}_{N+1*}$ will be a more efficient predictor. This is what one would expect under this situation.

It is of interest to consider the property of $\tilde{y}_{N+1*}$ when the model of Case 2 is correct i.e.

$$\tilde{y}_i = X\beta + X\tilde{\eta}_i + \tilde{e}_i \ , \ i = 1, 2, \ldots, N. \qquad (2.4.2)$$

In this model the degree of population and individual parameters are the same. The predictor $\tilde{y}_{N+1*}$ is no more unbiased and this introduces a bias of amount $\tilde{v}'\beta_2$. This will increase the $\text{MSE}(\tilde{y}_{N+1*})$ by an additional term $(\tilde{v}'\tilde{\beta}_2)^2$ in (2.3.4). However, the above conclusions are likely to hold even in this situation.

Suppose one is interested in comparison of MSE's for large value of $N$, in the neighbourhood of design points, as done by Allen (1971) and Hocking (1974). It is not difficult to see that the difference between average MSE's, average taken over all design points, is given by

$$\text{MSE}(\tilde{y}_{N+1*}) - \text{MSE}(\hat{y}_{N+1*}) = p^{-1} \left[ (q_1 - q)\sigma^2 + \sigma^4 \text{Tr}(X'_{2.1} X_{2.1} G^{-1}) \right.$$

$$\left. + \text{Tr} \{ F_{22}(X'_2 X_2 - X'_2 X_1 (X'_1 X_1)^{-1} X'_1 X_2) \} \right] ,$$

where

$$X_{2.1} = [I - X_1(X_1' \Omega X_1)^{-1} X_1' \Omega] X_2, \quad G = X_2' \Omega X_{2.1} = X_{2.1}' \Omega X_2,$$

and $\Omega = XFX' + \sigma^2 I$.

Again we see that when $F_{22}$ does not have small characteristic roots then the last two terms of the above expression are positive and large in comparison to the first term which is negative, thus making $\hat{y}_{N+1*}$ to be a better predictor.

## 2.5 EFFECT OF ESTIMATION OF PARAMETERS ON PREDICTORS

In this section we consider the problem of prediction for model (2.4.2). The MMSE predictor for a future observation, having structure

$$y_{N+1} = z_0' \beta + z_0' \eta_{N+1} + e_{N+1},$$

when $\sigma^2, \beta$ and F are known, is given by $z_0' \beta_{N+1}^b$, where

$$\hat{\beta}_{N+1}^b = \hat{\beta}_{N+1} - \sigma^2 U(F + \sigma^2 U)^{-1}(\hat{\beta}_{N+1} - \beta).$$

Here we shall consider a modified estimator when $\sigma^2, \beta$ and F are estimated from the previous observations $y_1, y_2, \ldots, y_N$. Reinsel (1984b) has observed that there is a considerable increase in variance when parameters are to be estimated from a sample. We modify predictor $\hat{y}_{N+1*}$, by introducing a constant $c$, as $z_0' \beta_{N+1}^e$ where

$$\beta_{N+1}^e = \hat{\beta}_{N+1} - cW_1 UB^{-1}(\hat{\beta}_{N+1} - \beta), \tag{2.5.1}$$

the superscript e denotes empirical Bayes type estimator.

Rao (1975) has obtained constant c so as to minimise $E \sum \{ \underset{\sim}{t}' (\underset{\sim}{\beta}_i^e - \underset{\sim}{\beta}_i) \}^2$ where summation extends over all N individuals. In the present case we find c so as to minimise the prediction error $E\{ Y_{N+1} - \hat{Y}_{N+1}(c) \}^2$ for any future observation where $\hat{Y}_{N+1}(c) = \underset{\sim}{z}_o' \underset{\sim}{\beta}_{N+1}^e$. It can be shown that

$$E\{ Y_{N+1} - \hat{Y}_{N+1}(c) \}^2 = \sigma^2 + \sigma^2 \underset{\sim}{z}_o' U \underset{\sim}{z}_o + N_3 (c^2 N_4 - 2cN) \sigma^4 \underset{\sim}{z}_o' U (F + \sigma^2 U)^{-1} U \underset{\sim}{z}_o$$

$$(2.5.2)$$

where

$$N_3 = (p-q)(N-q-2)^{-1}, \quad N_4 = (Np - Nq + 2)(N+1)(N-2)\{ (N-q-1)(N-q-4) \}^{-1}$$

Therefore, the optimum value of c that minimizes (2.5.2) is given by $c_o = N \cdot N_4^{-1}$. Putting this value of $c_o$ in (2.5.2) we obtain

$$E(Y_{N+1} - \hat{Y}_{N+1}(c))^2 = \sigma^2 + \sigma^2 \underset{\sim}{z}_o' U \underset{\sim}{z}_o - N^2 N_3 N_4^{-1} \sigma^4 \underset{\sim}{z}_o' U (F + \sigma^2 U)^{-1} U \underset{\sim}{z}_o.$$

$$(2.5.3)$$

The value given by Rao (1975) is $c_R = (N-q-2)$ $(Np - Nq + 2)^{-1}$ which differs considerably from $c_o$ for small values of N but for large N they are nearly equal.

Example :   For empirical comparison we consider the data provided by Strenio et al. (1983) (see Experiment 1 at the end of thesis) which is a modified version of Box (1950). We fit a first degree polynomial (q=2) for each rat. We omit the data on a rat and estimate the parameters from the observations at t = 1(1)4 on other rats. For different values of c we predict

Table 2.1:Observed and Predicted Weights (in gm) and Mean Square Deviation (MSD)at t=5  for Different Values of C

| Rat | Observed Weights | Predicted Weights | | | |
|-----|------------------|-------|----------|----------|-------|
|     |                  | 0 | $C_R$ | $C_O$ | $C^*$ |
| 1   | 176.00 | 158.50 | 158.64 | 158.57 | 158.76 |
| 2   | 174.00 | 180.00 | 173.08 | 176.80 | 167.69 |
| 3   | 201.00 | 179.50 | 170.37 | 175.27 | 163.26 |
| 4   | 157.00 | 159.00 | 156.36 | 157.78 | 154.30 |
| 5   | 148.00 | 145.00 | 148.69 | 146.71 | 151.56 |
| 6   | 152.00 | 166.00 | 153.73 | 160.32 | 144.18 |
| 7   | 156.00 | 126.50 | 139.24 | 132.40 | 149.15 |
| 8   | 154.00 | 151.50 | 164.97 | 157.74 | 175.45 |
| 9   | 138.00 | 137.00 | 140.49 | 138.62 | 143.21 |
| 10  | 177.00 | 165.00 | 168.61 | 166.67 | 171.41 |
| MSD |        | 203.50 | 172.22 | 172.33 | 240.76 |

the weight of omitted rat at p=5 given its weight at t=1(1)4.
This is done for each of the 10 rats. The predicted values
are then compared with known weight of the rat. The predictor
corresponding to c = 0 does not take into account the
information provided by the other individuals while the one
corresponding to $c* = (N+1)(N_p-N_q)^{-1}$ takes the estimated
values as real values. Table 2.1 contains observed and predicted
weights of rats at t = 5 and their mean square deviation
$(MSD) = \{y_i - \hat{y}_i(c)\}^2/10$, for c=0, $c_R$, $c_O$ and c*.

The prediction error calculated by using expression
(2.5.3) after substituting the estimated parameter values is
195.29, which compares reasonably well with the cross validation
value of 172.33.

The difference between MSD's for $c=c_R$ and $c_O$ is
negligible. However, using information on previous observations
makes errors much smaller as indicated by MSD's corresponding
to c=0 and $c=c_R$ (or $c_O$). The high MSD corresponding to c=c*
indicates that introduction of an optimum c is important when
N is moderate.

2.6   SELECTION OF DEGREE OF POLYNOMIAL

Let $y_{ti}$ be the measurement at time t (t = 1,2,...,p)
on individual i (i = 1,2,...,N). Assuming a polynomial RCR
model, we consider the selection of optimum degree of polynomial
for the purpose of prediction. Without any loss of generality
we can work with orthogonal polynomials. Suppose that response
$y_{ti}$ is expressed as

$$y_{ti} = \underset{\sim}{\Psi}'(t)\, \underset{\sim}{\beta}_i + e_{ti}, \quad t = 1,2,\ldots,p, \quad i = 1,2,\ldots,N \quad (2.6.1)$$

where $\underset{\sim}{\Psi}(t) = (\Psi_0(t), \Psi_1(t), \ldots, \Psi_s(t))'$ is a $(s+1) \times 1$ vector of coefficients of orthogonal polynomials. We assume that

$$E(e_{ti}) = 0, \quad \text{var}(e_{ti}) = \sigma^2, \quad \text{cov}(e_{ti}, e_{ui}) = 0 \text{ for } t \neq u,$$

$$\text{cov}(e_{ti}, e_{tj}) = 0 \text{ for } i \neq j, \quad E(\underset{\sim}{\beta}_i) = \underset{\sim}{\beta}, \quad \text{var}(\underset{\sim}{\beta}_i) = F,$$

and $\text{cov}(\underset{\sim}{\beta}_i, \underset{\sim}{e}_j) = 0$ for all $i$ and $j$,

where

$$\underset{\sim}{e}_j = (e_{i1}, e_{i2}, \ldots, e_{ip})'.$$

Under the model $(2.6.1)$, the least square estimator of $\underset{\sim}{\beta}_i$ is given by

$$\underset{\sim}{\beta}_i^{\ell} = \Psi'\, \underset{\sim}{y}_i$$

where

$\Psi = [\underset{\sim}{\Psi}(1), \underset{\sim}{\Psi}(2), \ldots, \underset{\sim}{\Psi}(p)]'$ is a $p \times (s+1)$ matrix. We know that when $\sigma^2, \underset{\sim}{\beta}$ and $F$ are known the optimum linear predictor of $g'\underset{\sim}{\beta}_i$ is $g'\underset{\sim}{\beta}_i^b$ where $\underset{\sim}{\beta}_i^b$ is given by the expression similar to $(2.3.1)$. When $\sigma^2, \underset{\sim}{\beta}$ and $F$ are unknown, Rao (1975) substitutes suitable constant multiplier of the following unbiased estimators, which are well known,

$$N\,\underset{\sim}{\beta}^* = \Sigma\,\underset{\sim}{\beta}_i^{\ell} \;;$$

$$N(p-s-1)\,\sigma^{*2} = \left[ \sum_1^N \sum_1^p y_{ti}^2 - \sum_1^N \sum_1^s (\beta_{ir}^{\ell})^2 \right];$$

$$(N-1)(F+\sigma^2 I)_* = \Sigma (\underset{\sim}{\beta}_i^{\ell} - \underset{\sim}{\beta}^*)(\underset{\sim}{\beta}_i^{\ell} - \underset{\sim}{\beta}^*)' = B_2;$$

in the expression for $\beta_{\sim i}^{b}$ to obtain the empirical Bayes estimator of $g'\beta_{\sim i}$ as $g'\beta_{\sim i}^{e}$. Here $\beta_{\sim i}^{e}$ is defined as

$$\beta_{\sim i}^{e} = \beta_{\sim i}^{\ell} - c(s)\ \sigma^{*2}\ B_2^{-1}\ (\beta_{\sim i}^{\ell} - \beta_{\sim i}^{*}) \qquad (2.6.2)$$

where, for any integer k,

$$c(k) = N(p-s-1)(N-k-3)(Np-Ns-N+2)^{-1}.$$

## 2.6.1 Choice of the degree of polynomial for extrapolation :

Here we consider the optimum choice of degree of polynomial for predicting $y_{p+1i}$ (i = 1,2,...,N) simultaneously. The criterion used is compound mean square prediction error (CMSPE), i.e. the total mean square error of prediction over all N individuals.

If we choose only the first (d+1) terms in (2.6.1), when in fact the true model has (s+1) terms, then the minimum MSE predictor of $y_{p+1i}$ is

$$y_{p+1i}^{e} = \psi'(p+1)\ \beta_{\sim i}^{e}\ ,$$

where $\psi_{\sim}(t) = (\psi_{\sim}(t)_1'\ ,\ \psi_{\sim}(t)_2')'$ and $\beta_{\sim i}^{e} = (\beta_{\sim 1i}^{e'}\ ,\ \beta_{\sim 2i}^{e'})'$ have partitions of dimensions d+1 and s-d, respectively. Using results of Rao (1975) it is not difficult to see that the optimum value of constant to be substituted in (2.6.2) is c(d) and

$$E\ \Sigma\ (\beta_{\sim 1i}^{e} - \beta_{\sim 1i})(\beta_{\sim 1i}^{e} - \beta_{\sim 1i})' = N\sigma^2 I - c(d)\ \sigma^4(F_{11} + \sigma^2 I)^{-1},$$

where $\beta_i = (\beta'_{1i}, \beta'_{2i})'$ are partitions of $\beta_i$ of dimensions $d+1$ and $s-d$, respectively, and $F$ is partitioned as

$$F = \begin{array}{c} d+1 \\ s-d \end{array} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}.$$
$$\quad\quad d+1 \quad\quad s-d$$

In this case the CMSPE is

CMSPE$(p+1)$

$$= N\sigma^2 + \Psi(p+1)'_1 \left[ N\sigma^2 I - c(d)\sigma^4(F_{11}+\sigma^2 I)^{-1} \right] \Psi(p+1)_1$$

$$+ N \Psi(p+1)'_2 \left[ F_{22}+\beta_2\beta'_2 \right] \Psi(p+1)_2$$

$$+ 2\Psi(p+1)'_1 \left[ c(d)\sigma^2(F_{11}+\sigma^2 I)^{-1}F_{12} \right] \Psi(p+1)_2$$

$$= N\sigma^2 + \Psi'(p+1) \begin{bmatrix} N\sigma^2 I - c(d)\sigma^4(F_{11}+\sigma^2 I)^{-1} & c(d)\sigma^2(F_{11}+\sigma^2 I)^{-1}F_{12} \\ c(d)\sigma^2 F_{21}(F_{11}+\sigma^2 I)^{-1} & N(F_{22}+\beta_2\beta'_2) \end{bmatrix} \Psi(p+1)$$

$$(2.6.3)$$

In practice the minimization of (2.6.3) over d cannot be carried out because $\sigma^2, \beta$ and $F$ are not known. We substitute unbiased estimators of quantities on right hand side of (2.6.3) to obtain an estimate of d.

Rao (1981) has considered this problem in the fixed coefficient setup by using least square estimators. The case of Rao is a particular case of this when $F$ is a null matrix and $c(d) = 0$. Under these conditions the expression (2.6.3) reduces to

$$N\sigma^2 + N\sigma^2 \, \underset{\sim}{\dot{\Psi}}(p+1)'_1 \, \underset{\sim}{\dot{\Psi}}(p+1)_1 + \sum_1^N \underset{\sim}{\dot{\Psi}}(p+1)'_2 \, \underset{\sim}{\beta}_{2i}$$

which agrees with equation (4.1.2) of Rao.

2.6.2  <u>Choice of d for interpolation</u> :

Suppose our aim is to predict the value of response in the neighbourhood of time points  considered in the design. In such a case it may be appropriate to average the CMSPE over all design points as has been done by Allen (1971) and Hocking (1974). Since $\underset{\sim}{\Psi}(t)$'s are coefficients of orthogonal polynomials it is not difficult to observe from (2.6.3) that the average CMSPE, average taken over all p points, is

$$\text{Average CMSPE} = p^{-1}\left[N(p+d+1)\sigma^2 - c(d)\sigma^4 \, \text{Tr}(F_{11}+\sigma^2 I)^{-1} + N(F_{22}+\underset{\sim}{\beta}_2\underset{\sim}{\beta}'_2)\right]$$

$$(2.6.4)$$

Again, in the absence of known values of parameters we substitute unknown parameters by their unbiased estimators. We obtain values of d by numerically minimizing (2.6.4).

2.6.3  <u>Example</u> :   For the purpose of illustration we use the data on weights of 13 male mice measured at intervals of 3 days over 21 days as reported by Williams and Izenman (1981) (see Experiment 2). We estimate the parameters by taking s = 4 and using data upto 18 days. The values of estimated CMSPE obtained by using (2.6.3) for extrapolation purposes at day 21 are reported in column (2) of Table 2.2. The estimated average CMSPE, to be used for the purpose of interpolation, obtained from (2.6.4) are

Table 2.2:Compound Mean Square Prediction Error(CMSPE)

for Polynomials of Different Degrees

| Degree of the Polynomial (1) | CMPSE at t=5 (2) | Average CMPSE (3) | SSD (4) |
|---|---|---|---|
| 0 | 0.7929 | 0.7336 | 1.8042 |
| 1 | 0.9170 | 0.0215 | 0.2169 |
| 2 | 0.3659 | 0.0055 | 0.0846* |
| 3 | 0.1317* | 0.0048* | 0.1267 |
| 4 | 0.3429 | 0.0054 | 0.4355 |

* denotes the minimum value.

presented in column (3) of Table 2.2. The sum of square of deviations (SSD) of predictor based on (2.6.2) for day 21 and its observed value are reported in column (4) of Table 2.2.

If we had data upto 18 days and wanted to predict the weights on 21st day then, looking at the values of CMSPE in column (2) of Table 2.2, we would have fitted a third degree polynomial. However, from column (4) of Table 2.2, we see that a quadratic is a better fit than a cubic. This small disagreemer' may arise due to reasons such as inadequacy of the fit of RCR model and small amount of data available. **For** interpolation purposes a cubic model should be fitted.

## 2.7 SELECTION OF COVARIATES

In this section we consider the selection of covariates for covariance adjustment which are optimum in some sense. We follow procedure similar to that of Rao (1967). Rao's (1967) procedure is as follows.

Let $\underset{\sim}{T}_1$ and $\underset{\sim}{T}_2$ be two vector statistics of order m and r such that $E(\underset{\sim}{T}_1) = \underset{\sim}{\tau}$ and $E(\underset{\sim}{T}_2) = \underset{\sim}{0}$ where $\underset{\sim}{\tau}$ is a m vector of unknown parameters. If $cov(\underset{\sim}{T}_1, \underset{\sim}{T}_2) \neq 0$ then an estimator of which is better than $\underset{\sim}{T}_1$ can be obtained. Let

$$var(\underset{\sim}{T}) = var \begin{bmatrix} \underset{\sim}{T}_1 \\ \underset{\sim}{T}_2 \end{bmatrix} = \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} .$$

If we consider an estimator of $\underset{\sim}{\tau}$ as

$$\underset{\sim}{\tau^*} = \underset{\sim 1}{T} - \Lambda_{12} \Lambda_{22}^{-1} \underset{\sim 2}{T}$$

then it is not difficult to show that $\underset{\sim}{\tau^*}$ has smaller variance than $\underset{\sim 1}{T}$. If $\Lambda$ is not known but an estimate of $\Lambda$, given by U,

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

is available, then it is natural to consider an estimator of as

$$\underset{\sim}{\hat{\tau}} = \underset{\sim 1}{T} - U_{12} U_{22}^{-1} \underset{\sim 2}{T}.$$

Rao (1967) has shown that if

$$\underset{\sim}{T} \sim N_{m+r} \left( \begin{bmatrix} \underset{\sim}{\tau} \\ \underset{\sim}{0} \end{bmatrix}, \Lambda \right)$$

and

$$U \sim W(\Lambda, f, m+r)$$

then

$$E(\underset{\sim}{\hat{\tau}}) = \underset{\sim}{\tau} ; \quad var(\underset{\sim}{\hat{\tau}}) = \frac{f-1}{f-r-1} \left[ \Lambda_{11} - \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21} \right]. \quad (2.7.1)$$

The estimator $\underset{\sim}{\hat{\tau}}$ may not be necessarily better than $\underset{\sim 1}{T}$. Infact $\underset{\sim 1}{T}$ is likely to be better than $\underset{\sim}{\hat{\tau}}$ when $\Lambda_{12}$ is small and r is large. In such cases it may be advantageous to choose a subset of $\underset{\sim 2}{T}$ which may be highly correlated with $\underset{\sim 1}{T}$. For a model with

$$E(\underset{\sim}{y}) = X\underset{\sim}{\tau} ; \quad var(\underset{\sim}{y}) = X\Lambda X' + \sigma^2 I = Q, \text{ say,}$$

where $\underset{\sim}{y}$ is an observable p component random vector and X is a known p×m known design matrix, Rao (1967) recommends the use of

characteristic vectors corresponding to the dominant characteristi
roots of $[I - X(X'X)^{-1}X']$ U as covariates, where U is the
estimated dispersion matrix of $\underset{\sim}{T}$, given by

$$fU = \begin{bmatrix} (X'X)^{-1} X'SX(X'X)^{-1} & (X'X)^{-1}X'SZ \\ Z'SX(X'X)^{-1} & Z'SZ \end{bmatrix}, \quad (2.7.2)$$

where Z is a p x r matrix of rank r such that $Z'X = 0$, and

$$S \sim W(\Omega; f, p).$$

This, however, is not based on any optimality criterion. One
may seek the choice of covariates based on trace minimization
and generalised variance minimization criteria which are
parallel to A-optimality and D-optimality criteria in design
of experiments.

## 2.7.1 Choice of covariates for trace minimization criterion :

Suppose one is interested in selecting a linear combinatio
of components of $\underset{\sim}{T_2}$ so that the trace of the dispersion matrix
of covariance adjusted estimator of $\underset{\sim}{\tau}$ is minimum. Let $\underset{\sim 1}{\ell'} \underset{\sim 2}{T}$
be such a linear combination. Then

$$\begin{bmatrix} \underset{\sim 1}{T} \\ \underset{\sim 1}{\ell'} \underset{\sim 2}{T} \end{bmatrix} \sim N_{m+1} \left( \begin{bmatrix} \underset{\sim}{\tau} \\ \underset{\sim}{0} \end{bmatrix}, \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \underset{\sim 1}{\ell} \\ \underset{\sim 1}{\ell'} \Lambda_{21} & \underset{\sim 1}{\ell'} \Lambda_{22} \underset{\sim 1}{\ell} \end{bmatrix} \right).$$

We define a covariance adjusted estimator of $\underset{\sim}{\tau}$ as

$$\underset{\sim 1}{\hat{\tau}} = \underset{\sim 1}{T} - U_{12} \underset{\sim 1}{\ell} (\underset{\sim 1}{\ell'} U_{22} \underset{\sim 1}{\ell})^{-1} \underset{\sim 1}{\ell'} \underset{\sim 2}{T}.$$

From expression similar to $(2.7.1)$ it follows that

$$E(\hat{\tau}_1) = \underset{\sim}{\tau} \; ; \; var(\hat{\tau}_1) = \frac{f-1}{f-2} \; [\Lambda_{11} - \Lambda_{12} \; \underset{\sim 1}{\ell}(\underset{\sim 1}{\ell'} \; \Lambda_{22} \; \underset{\sim 1}{\ell})^{-1} \; \underset{\sim 1}{\ell'} \; \Lambda_{21}] \; .$$

For minimizing $\mathrm{Tr} \; [var(\hat{\tau}_1)]$ with respect to $\underset{\sim 1}{\ell}$, it is not difficult to see that the best strategy is to choose $\underset{\sim 1}{\ell}$ as the vector corresponding to the largest root $\varphi_1$ satisfying

$$(\Lambda_{21} \; \Lambda_{12} - \varphi_1 \; \Lambda_{22}) \; \underset{\sim}{\ell} = \underset{\sim}{0}.$$

In the absence of the knowledge of $\Lambda$ we substitute its estimate U.

## 2.7.2  Choice of covariates for generalized variance minimization criterion :

In this subsection we look for the best strategy in terms of minimum generalised variance of covariance adjusted estimator, say $\underset{\sim 2}{\tau}$, of $\underset{\sim}{\tau}$ . Let

$$\hat{\tau}_2 = \underset{\sim 1}{T} - U_{12} \; \underset{\sim 2}{\ell} \; (\underset{\sim 2}{\ell'} \; U_{22} \; \underset{\sim 2}{\ell})^{-1} \; \underset{\sim 2}{\ell'} \; \underset{\sim 2}{T}$$

be such a covariance adjusted estimator.  It follows that

$$E(\hat{\tau}_2) = \underset{\sim}{\tau} \; , \; var(\hat{\tau}_2) = \frac{f-1}{f-2} \; [\Lambda_{11} - \Lambda_{12} \; \underset{\sim 2}{\ell'}(\underset{\sim 2}{\ell'} \; \Lambda_{22} \; \underset{\sim 2}{\ell})^{-1} \; \underset{\sim 2}{\ell'} \; \Lambda_{21}] \; .$$

By observing that $|var(\hat{\tau}_2)|$ can be written as

$$|var(\hat{\tau}_2)| = (\frac{f-1}{f-2})^m . \; |\Lambda_{11}| \; (1 - \frac{\underset{\sim 2}{\ell'} \; \Lambda_{21} \; \Lambda_{11}^{-1} \; \Lambda_{12} \; \underset{\sim 2}{\ell}}{\underset{\sim 2}{\ell'} \; \Lambda_{22} \; \underset{\sim 2}{\ell}})$$

it is not difficult to see that the best strategy is to take $\underset{\sim 2}{\ell}$ as the vector corresponding to the largest root $\varphi_2$ satisfying

$$(\Lambda_{21} \; \Lambda_{11}^{-1} \; \Lambda_{12} - \varphi_2 \; \Lambda_{22}) \; \underset{\sim}{\ell} = \underset{\sim}{0}.$$

Again, in the absence of the knowledge of $\Lambda$, we substitute its estimate U. Here $\varphi_2$ is nothing but the square of canonical correlation between $T_{\sim 1}$ and $T_{\sim 2}$.

In both cases discussed above (sec. 2.7.1 and 2.7.2) more covariates can be chosen by considering the vectors corresponding to the next largest roots.

2.7.3 <u>Example</u> : For the purpose of illustration we use the data of ramus height, measured in mm, on 20 individuals over 2 years at half yearly intervals, as reported by Elston and Grizzle (1962) (see Experiment 3). A linear curve (m = 2) is fitted to the data using coefficients of orthogonal polynomials. The matrix S is taken as the matrix of sum of squares and cross-products corrected for mean with 19 (=f) degrees of freedom. Here r is taken as 2. The matrix U can be obtained by using (2.7.2) where X and Z are suitable matrices formed by using coefficients of orthogonal polynomials.

Table 2.3 summarizes the results corresponding to three methods of choice of covariates, viz., Rao's(R), trace minimization(a) and generalised variance minimization(b). It may be noted that the results for no covariates are the same by these three methods, as are for two (all) covariates, as they should be. The methods (a) and (b) are the best of the three, in their respective sense, when one covariate is used. However, there is a slight increase in the average variance with the use of covariates for these data. This shows that the correlations

Table 2.3:Average and Generalised Variances by Three Methods of Choice of Covariates viz.,(i)Rao's,(ii)Trace Minimization(a) and (iii) Generalised Variance Minimization(b) For Ramus Height Data

| | no covariate | one covariate | | | two covariates |
| --- | --- | --- | --- | --- | --- |
| | | (R) | (a) | (b) | (b) |
| $\hat{\tau}_2$ | $\begin{bmatrix} 50.075 \\ 0.467 \end{bmatrix}$ | $\begin{bmatrix} 50.072 \\ 0.463 \end{bmatrix}$ | $\begin{bmatrix} 50.053 \\ 0.458 \end{bmatrix}$ | $\begin{bmatrix} 50.074 \\ 0.465 \end{bmatrix}$ | $\begin{bmatrix} 50.050 \\ 0.465 \end{bmatrix}$ |
| $20\times \mathrm{var}(\hat{\tau}_2)$ | $\begin{bmatrix} 6.265 & 0.090 \\ 0.090 & 0.085 \end{bmatrix}$ | $\begin{bmatrix} 6.610 & 0.068 \\ 0.068 & 0.058 \end{bmatrix}$ | $\begin{bmatrix} 6.556 & 0.065 \\ 0.065 & 0.078 \end{bmatrix}$ | $\begin{bmatrix} 6.617 & 0.072 \\ 0.072 & 0.058 \end{bmatrix}$ | $\begin{bmatrix} 7.427 & 0.082 \\ 0.082 & 0.065 \end{bmatrix}$ |
| $20\times \mathrm{Tr}[\mathrm{var}(\hat{\tau}_2)]$ | 6.350 | 6.668 | 6.634 | 6.675 | 7.493 |
| $400\times |\mathrm{var}(\hat{\tau}_2)|$ | 0.524 | 0.380 | 0.506 | 0.377 | 0.480 |

are not large enough so as to justify the use of the covariates,
as for as the average variances are concerned.

## 2.8  DISCUSSION AND SUMMARY

In most of the work on the RCR models reported in the
literature, it is assumed that the dimensions of population
and individual regression parameters are the same.  But this
may not always be true.  Fearn (1977) has considered a model
wherein the dimension of the population regression parameters
is smaller than for individual's.  It is possible that the
dimension of the population regression parameters is smaller
than for individual's but the analyst may wrongly proceed to
analyse by assuming that they are the same.  In case 1 we have
studied the effect on prediction error when the dimension of
individual's parameters is larger than what has been considered
for analyses purposes.  In case 2 the effect on prediction error
has been studied when the dimension of population parameters
is considered to be larger than the actual one.  The judgement
regarding the relative efficiencies in these two situations
should be based on $\Gamma_{2.1}$, the matrix of regression coefficients
of $\hat{\beta}_{2i}$ on $\hat{\beta}_{1i}$, and $F_{22}$.  On the basis of expression (2.4.1) we
can say that if $F_{22}$ has large characteristic roots then the MSE
for case 2 is likely to be smaller than for the case 1.

The problem of prediction for a new individual is
considered in Section 2.5.  Rao (1975) has obtained the value of
a shrinkage parameter c to be used when estimated values are

substituted in place of the parameters, that minimizes $E \sum \{ t'_\sim (\beta^c_{\sim i} - \beta_{\sim i}) \}^2$, where the summation extends over all individual. A constant multiplier c which minimizes $E [y_{N+1} - \hat{y}_{N+1}(c)]^2$ for a new individual has been obtained. It can be seen that these two values differ considerably for small N but for large N they are nearly equal, as expected.

The assumption regarding the knowledge about the degree of the polynomial is seldom likely to hold in practice. A more realistic situation is one in which the analyst has to decide the degree based on the data. This problem in the setup of RCR model has been considered in Section 2.6. The choice of the degree can be made on the basis of values of estimated CMSPE [equation (2.6.3)] or average CMSPE [equation (2.6.4)] as the case may be.

Another important aspect in the analysis of polynomial models is the selection of covariates for covariance adjustment to obtain more efficient estimators. This problem has been considered by Rao (1965,1967) and Grizzle and Allen (1969). Rao (1967) has suggested a method which is not based on any optimality criterion. Two methods are proposed in Section 2.7 which provide the best strategy for the choice of covariates in terms of minimizing the average variance and the generalised variance , respectively, of the covariance adjusted estimator.

CHAPTER III

ESTIMATION AND PREDICTION IN A MULTIVARIATE RANDOM COEFFICIENT
REGRESSION MODEL WHEN DESIGNS ARE DIFFERENT FOR DIFFERENT
CHARACTERISTICS

## 3.1 INTRODUCTION

The RCR model proposed by Rao (1965) has received wide
attention of the researchers working in the area of growth
curve analysis. Rao (1965,1967,1975), Fisk (1967) and Swamy
(1971) have laid a strong theoretical foundation for studies
in RCR models. Since then, this model has been generalised in
various directions. For example, Strenio et al. (1983) extend
the work by introducing covariates in the model. In a series
of papers Reinsel (1982,1984a,1985) generalises the growth
curve and the RCR models to a multivariate case. He imposes
certain restrictions on the model such as the response for
each characteristic be approximated by polynomials of the same
degree and the set of covariates be common for all the
characteristics. This leads to a considerable simplification
in the results. However, there are situations where it may
not be appropriate to assume the same degree of polynomial for
all the characteristics. One such situation has been described
in the following:

Let us consider a problem of predicting the yield of a
crop which is available at the time of harvest. The records on

growth characters, such as height, diameter and leaf width, may
be available at various time points. There is no reason why
polynomials of the same degree should be good fit to all the
characteristics. Moreover, some variates that may be covariates
for height may not be good covariates for, say, leaf width. Thus
there is a need to generalise the model of Reinsel (1984a) to
allow for different degrees of polynomial for different
characteristics.

In Section 3.2 we propose a generalised multivariate RCR
model and obtain the maximum likelihood (ML) estimators, which
are iterative, for the parameters of the model. In Section 3.3
expression for approximate mean square error (MSE) of the
predictors, based on the ML estimators, has been derived. A
non iterative estimation procedure for a particular case of the
model, which may provide initial values for the ML procedure,
is suggested in Section 3.4. A Monte Carlo simulation study
to compare the performance of the predictors based on the
ML estimators and some noniterative estimators is described
in Section 3.5. The results of this study are discussed in
Section 3.6. In Section 3.7 the results of a worked out
example are given. Section 3.8 contains discussion and
summary of the work done in this chapter.

## 3.2 MODEL AND ESTIMATION OF PARAMETERS

Let us assume that there are N individuals, and m
characteristics are observed on each individual. The

jth $(j = 1, 2, \ldots, m)$ characteristic is observed at $p_j$ occasions and its response can be approximated by a polynomial of $(q_j - 1)$ degree in time. Let

$$\underset{\sim}{y}_{ij} = A_j \underset{\sim}{\pi}_{ij} + \underset{\sim}{e}_{ij} \, , \quad i = 1, \ldots, N; \quad j = 1, \ldots, m; \quad (3.2.1)$$

where $\underset{\sim}{y}_{ij}$ is an observable $p_j$-vector random variable corresponding to the observations on the j-th characteristic of the i-th individual; $A_j$ is a known $p_j \times q_j$ design matrix, and $\underset{\sim}{\pi}_{ij}$ is a $q_j \times 1$ column vector of random regression coefficients corresponding to the j-th characteristic of the i-th individual. We assume that

$$E(\underset{\sim}{e}_{ij} | \underset{\sim}{\pi}_{ij}) = \underset{\sim}{0} \, ; \quad \text{var}(\underset{\sim}{e}_{ij} | \underset{\sim}{\pi}_{ij}) = \sigma_j^2 I.$$

We shall view $\underset{\sim}{\pi}_{ij}$ as a random parameter vector related to covariates as

$$\underset{\sim}{\pi}_{ij} = X_{ij} \underset{\sim}{\gamma}_j + \underset{\sim}{\eta}_{ij} \, , \quad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, m. \quad (3.2.2)$$

where $X_{ij}$ is a $q_j \times d_j$ matrix of known covariates and $\underset{\sim}{\gamma}_j$ is a $d_j$-vector of fixed parameters.

The equations $(3.2.1)$ and $(3.2.2)$ can be written as

$$\underset{\sim}{y}_i = A\underset{\sim}{\pi}_i + \underset{\sim}{e}_i \, ; \quad \underset{\sim}{\pi}_i = X_i \underset{\sim}{\gamma} + \underset{\sim}{\eta}_i \, ; \quad i = 1, \ldots, N; \quad (3.2.3)$$

where

$$\underset{\sim}{y}_i = (\underset{\sim}{y}'_{i1}, \ldots, \underset{\sim}{y}'_{im})' \, ; \quad A = \text{diag}(A_1, \ldots, A_m); \quad \underset{\sim}{\pi}_i = (\underset{\sim}{\pi}'_{i1}, \ldots, \underset{\sim}{\pi}'_{im})' \, ;$$

$$X_i = \text{diag}(X_{i1}, \ldots, X_{im}); \underset{\sim}{\eta} = (\underset{\sim}{\eta}'_1, \ldots, \underset{\sim}{\eta}'_m)' \, ; \quad \underset{\sim}{e}_i = (\underset{\sim}{e}'_{i1}, \ldots, \underset{\sim}{e}'_{im})' \, ;$$

$$\underset{\sim}{\eta}_i = (\underset{\sim}{\eta}'_{i1}, \ldots, \underset{\sim}{\eta}'_{im})'$$

and

$$E(\underset{\sim}{e}_i | \underset{\sim}{\pi}_i) = \underset{\sim}{0} \; ; \; \operatorname{var}(\underset{\sim}{e}_i | \underset{\sim}{\pi}_i) = V = \operatorname{diag}(\sigma_1^2 I_{p_1}, \ldots, \sigma_m^2 I_{p_m});$$

$$E(\underset{\sim}{\eta}_i | \underset{\sim}{\gamma}) = \underset{\sim}{0} \; ; \; \operatorname{var}(\underset{\sim}{\eta}_i | \underset{\sim}{\gamma}) = F.$$

Note that $\underset{\sim}{y}_i$ and $\underset{\sim}{\pi}_i$ are vectors of order $p(= \Sigma\, p_j)$ and $q(= \Sigma\, q_j)$, respectively.

Now we may write (3.2.3) as

$$\underset{\sim}{y}_i = A x_i \underset{\sim}{\gamma} + \underset{\sim}{u}_i$$

where

$$\underset{\sim}{u}_i = A\, \underset{\sim}{\eta}_i + \underset{\sim}{e}_i$$

with

$$E(\underset{\sim}{u}_i) = \underset{\sim}{0} \; ; \; \operatorname{var}(\underset{\sim}{u}_i) = AFA' + V = \Omega \; , \; \text{say.}$$

When $\underset{\sim}{u}_i$'s are jointly normally distributed, the likelihood function is

$$L(\underset{\sim}{\gamma}, \sigma_j^2, F | \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N) \propto |\Omega|^{-\frac{1}{2}N} . \exp(-\tfrac{1}{2} \Sigma\, \underset{\sim}{u}'_i\, \Omega^{-1}\, \underset{\sim}{u}_i),$$

where $\underset{\sim}{u}_i = \underset{\sim}{y}_i - A X_i \underset{\sim}{\gamma}$

After some matrix manipulation this may be written as

$$2 \log L(\underset{\sim}{\gamma}, \sigma^2, A | \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N)$$

$$= \text{const.} - N \log|V| - N \log |\Lambda| + N \log |U| - \Sigma\, \underset{\sim}{u}'_i V^{-1} \underset{\sim}{u}_i + \Sigma\, \underset{\sim}{u}'_i V^{-1} AUA' V^{-1}\, \underset{\sim}{u}_i$$

$$- \Sigma\, \underset{\sim}{u}'_i V^{-1} AU\Lambda^{-1} UA' V^{-1} \underset{\sim}{u}_i.$$

where $U = (A'V^{-1}A)^{-1}$ ; $\Lambda = F + U.$

The maximum likelihood estimators of $\underset{\sim}{\gamma}$, $\Lambda$ and $\sigma_j^2$ are obtained as

$$\hat{\underset{\sim}{\gamma}} = (\Sigma X_i' \hat{\Lambda}^{-1} X_i)^{-1} \Sigma X_i' \hat{\Lambda}^{-1} \hat{\underset{\sim}{\pi}}_i ; \qquad (3.2.4)$$

$$\hat{\Lambda} = N^{-1}.UA'V^{-1}(\Sigma \underset{\sim}{r}_i \underset{\sim}{r}_i')V^{-1}AU ; \qquad (3.2.5)$$

$$\hat{\sigma}_j^2 = \{N(p_j - q_j)\}^{-1}.Tr[S_j\{I - A_j(A_j'A_j)^{-1}A_j'\}] , 1 \le j \le m; \qquad (3.2.6)$$

where

$$\hat{\underset{\sim}{\pi}}_i = (A'A)^{-1}A' \underset{\sim}{y}_i ; \quad S_j = \Sigma \underset{\sim}{r}_{ij} \underset{\sim}{r}_{ij}' ; \quad \underset{\sim}{r}_{ij} = \underset{\sim}{y}_{ij} - A_j X_{ij} \hat{\underset{\sim}{\gamma}}_j ;$$

$$\underset{\sim}{r}_i = \underset{\sim}{y}_i - A X_i \hat{\underset{\sim}{\gamma}} .$$

An estimator of F may be obtained as $\hat{F} = \hat{\Lambda} - \hat{U}.$ For the sake of simplicity we have maximized the likelihood with respect to parameters $\underset{\sim}{\gamma}$, $\sigma_j^2$ and $\Lambda.$ Consequently, this does not give ML estimators of the original parameters and, in fact, estimate of F may fail to be positive definite in some cases. Rao (1975) and Reinsel (1984a) have also used these estimators and they are noniterative in their situations.

The equations (3.2.4) and (3.2.5) are solved iteratively till the convergence is achieved. Note that the right hand side of (3.2.6) can be written as

$$\{N(p_j - q_j)\}^{-1} \Sigma(\underset{\sim}{y}_{ij} - A_j \hat{\underset{\sim}{\pi}}_{ij})'(\underset{\sim}{y}_{ij} - A_j \hat{\underset{\sim}{\pi}}_{ij})$$

and the estimator $\hat{\sigma}_j^2$ of $\sigma_j^2$ is noniterative.

A particular case of interest, which introduce a considerable simplification, is when all $A_j$'s are same and

all the characteristics have common covariates, as considered
by Reinsel (1984a). For this case the equations (3.2.4) and
(3.2.5) are noniterative. The estimator of $\underset{\sim}{\gamma}$ is given by

$$\hat{\underset{\sim}{\gamma}} = (\Sigma X_i' X_i)^{-1} \; \Sigma X_i' \; \hat{\underset{\sim}{\pi}}_i.$$

The estimators are the same as obtained by Reinsel (1984a).

Since the variance component estimates obtained by the
ML method are even and translation invariant functions of
observations, it can be seen that $\hat{\underset{\sim}{\gamma}}$ is an unbiased estimator of
$\underset{\sim}{\gamma}$ (Kackar and Harville, 1981). Its asymptotic variance is
$(\Sigma X_i' \Lambda^{-1} X_i)^{-1}$. It is not difficult to see that $\hat{\Lambda}$ is a consistent
estimator of $\Lambda$. Moreover, $N(p_j-q_j)\hat{\sigma}_j^2 \sim \sigma_j^2 X^2$ on $N(p_j-q_j)$ degrees
of freedom.

## 3.3 PREDICTORS AND THEIR ERRORS

It may be of considerable interest to predict the $s_2$
components $\underset{\sim}{y}_{o2}$ of $\underset{\sim}{y}_o$ for a new individual, given its $s_1$
components $\underset{\sim}{y}_{o1}$ and information on auxiliary variable $X_o$. A
similar problem of prediction is considered by Fearn (1975),
Lee and Geisser (1975), Rao (1975) and Reinsel (1984 a,b)
and others.

The minimum mean square error predictor of $\underset{\sim}{y}_{o2}$, given
$\underset{\sim}{y}_{o1}$, is given by

$$\hat{\underset{\sim}{y}}_{o2} = A_{o2} X_o \underset{\sim}{\gamma} + A_{o2} F A_{o1}' (A_{o1} F A_{o1}' + V_1)^{-1} (\underset{\sim}{y}_{o1} - A_{o1} X_o \underset{\sim}{\gamma}); \qquad (3.3.1)$$

where V and A are partitioned as $V = diag(V_1, V_2)$ and

$A = (A'_{o1} : A'_{o2})'$, corresponding to partitions $\underset{\sim}{y}_{o1}$ and $\underset{\sim}{y}_{o2}$ of $\underset{\sim}{y}_o$, respectively. The mean square prediction error (MSPE) matrix of (3.3.1) is given by

$$E\left[(\underset{\sim}{y}_{o2} - \hat{\underset{\sim}{y}}_{o2})(\underset{\sim}{y}_{o2} - \hat{\underset{\sim}{y}}_{o2})'\right]$$

$$= (A_{o2} F A'_{o2} + V_2) - A_{o2} F A'_{o1} (A_{o1} F A'_{o1} + V_1)^{-1} A_{o1} F A'_{o2} = \Omega_{2.1}. \qquad (3.3.2)$$

Using some matrix identities, the equation (3.3.1) can be written as

$$\hat{\underset{\sim}{y}}_{o2} = A_{o2} \left[\hat{\underset{\sim}{\pi}}_o - U_1 (F + U_1)^{-1} (\hat{\underset{\sim}{\pi}}_o - X_o \underset{\sim}{\gamma})\right], \qquad (3.3.3)$$

where

$$\hat{\underset{\sim}{\pi}}_o = (A'_{o1} A_{o1})^{-1} A'_{o1} \underset{\sim}{y}_{o1} \; ; \; U_1 = (A'_{o1} V_1^{-1} A_{o1})^{-1} \; ;$$

provided all the necessary inverses exist. A particular case of great interest, where $A'_{o1} A_{o1}$ and $A'_{o1} V_1^{-1} A_{o1}$ may not be invertible, is discussed in details in subsection 3.3.1.

Now suppose that $\sigma_j^2$ ($j = 1, 2, \ldots, m$) and $F$ are known and we are required to estimate $\underset{\sim}{\gamma}$ from the sample. In this case a natural predictor of $\underset{\sim}{y}_{o2}$ is obtained by substituting $\hat{\underset{\sim}{\gamma}}$ in place of $\underset{\sim}{\gamma}$ in (3.3.3), i.e.

$$\tilde{\underset{\sim}{y}}_{o2} = A_{o2} \left[\hat{\underset{\sim}{\pi}}_o - U_1 (F + U_1)^{-1} (\hat{\underset{\sim}{\pi}}_o - X_o \hat{\underset{\sim}{\gamma}})\right]. \qquad (3.3.4)$$

Prediction error increases due to uncertainty about $\underset{\sim}{\gamma}$ by a quantity

$$E\{(\hat{\underset{\sim}{y}}_{o2} - \tilde{\underset{\sim}{y}}_{o2})(\hat{\underset{\sim}{y}}_{o2} - \tilde{\underset{\sim}{y}}_{o2})'\} = H\{var(\hat{\underset{\sim}{\gamma}})\}H',$$

where

$$H = A_{o2} U_1 (F+U_1)^{-1} X_o \; .$$

Thus the MSE matrix of (3.3.4) is

$$E\{ (\underset{\sim}{y}_{o2} - \underset{\sim}{\tilde{y}}_2)(\underset{\sim}{y}_{o2} - \underset{\sim}{\tilde{y}}_2)' \} = \Omega_{2.1} + H\{ \text{var}(\underset{\sim}{\hat{\gamma}}) \} H' , \qquad (3.3.5)$$

The predictor of $\underset{\sim}{y}_{o2}$ , when all the parameters have to be estimated from the sample, is given by

$$\underset{\sim}{y}_{o2}^{*} = A_{o2}\{ \underset{\sim}{\hat{\pi}}_o - \hat{U}_1 (\hat{F}+\hat{U}_1)^{-1}(\underset{\sim}{\hat{\pi}}_o - X_o\underset{\sim}{\hat{\gamma}}) \} \; . \qquad (3\ 3.6)$$

An estimator of $F+U_1$ is given by $\hat{\Lambda} - \hat{U}+\hat{U}_1$ , where $\hat{U}$ and $\hat{U}_1$ are obtained by substituting $\hat{\sigma}_j^2$ in place of $\sigma_j^2$ in $U$ and $U_1$, respectively. It is not difficult to see that $\hat{F}+\hat{U}_1$ so obtained is always positive definite. This gives

$$\underset{\sim}{\tilde{y}}_{o2} - \underset{\sim}{y}_{o2}^{*} = A_{o2}\{ \hat{U}_1 (\hat{F}+\hat{U}_1)^{-1} - U_1 (F+U_1)^{-1} \} (\underset{\sim}{\hat{\pi}}_o - X_o\underset{\sim}{\hat{\gamma}}) \; .$$

Using the approximation of Appendix 3.1 it is not difficult to observe that

$$E\{ (\underset{\sim}{y}_{o2} - \underset{\sim}{\tilde{y}}_{o2})(\underset{\sim}{\tilde{y}}_{o2} - \underset{\sim}{y}_{o2}^{*})' \} = O,$$

when N is large. Moreover (see Appendix 3.1)

$$E\{ (\underset{\sim}{\tilde{y}}_{o2} - \underset{\sim}{y}_{o2}^{*})(\underset{\sim}{\tilde{y}}_{o2} - \underset{\sim}{y}_{o2}^{*})' \}$$

$$\simeq A_{o2}[N^{-1}\{ U_1 (F+U_1)^{-1}(F+U)Q(F+U)(F+U_1)^{-1}U_1$$

$$+ \text{Tr}\{ (F+U)Q\} U_1 (F+U_1)^{-1}(F+U)(F+U_1)^{-1}U_1 \} + P + U_1 (F+U_1)^{-1}(T-P)$$

$$+ (T-P)(F+U_1)^{-1}U_1 + U_1 (F+U_1)^{-1}D(F+U_1)^{-1}U_1] A_{o2}' ; \qquad (3.3.7)$$

where $Q, P, T$ and $D$ are defined in the Appendix 3.1. This gives the total mean square error of (3.3.6) as

$$E\{ (\underset{\sim}{y}_{o2} - \underset{\sim}{y}_{o2}'')(\underset{\sim}{y}_{o2} - \underset{\sim}{y}_{o2}^*) \}$$

$$\simeq \Omega_{2.1} + H\{ var(\hat{\underset{\sim}{\gamma}}) \} H' + A_{o2} [ N^{-1} \{ U_1(F+U_1)^{-1}(F+U)Q(F+U)(F+U_1)^{-1}U_1$$

$$+ Tr\{(F+U)Q\} U_1(F+U_1)^{-1}(F+U)(F+U_1)^{-1}U_1 \} + P + U_1(F+U_1)^{-1}(T-P)$$

$$+ (T-P)(F+U_1)^{-1}U_1 + U_1(F+U_1)^{-1}D(F+U_1)^{-1}U_1 ] A_{o2}' . \qquad (3.3.8)$$

3.3.1 <u>An application to preharvest forecasting</u> : Consider the problem of preharvest forecasting. Measurements on biometric characters may be available over a period of time but the yield is available only at the time of harvest. Since yield is one of the m characteristics, the matrix $A_{o1}$ is of the form $A_{o1} = (A_{11}|0)$ and so $A_{o1}'A_{o1}$ is not invertible. The derivation of prediction variance when $A_{o1}'A_{o1}$ is not invertible but $A_{11}'A_{11}$ is invertible is considered in the following :

Without any loss of generality, let us consider the case when $s_2 = 1$ i.e. $y_{o2}$ is a scalar and $A_{o2} = \underset{\sim}{a}_2' = (\underset{\sim}{0}', 1)$ . The minimum mean square error predictor $(\hat{y}_{o2})$ of $y_{o2}$, given $\underset{\sim}{y}_{o1}$, and its variance are given by expressions (3.3.1) and (3.3.2), respectively. The predictor of $y_{o2}$ when only $\underset{\sim}{\gamma}$ has to be estimated from the sample is given by

$$\underset{\sim}{\tilde{y}}_{o2} = A_{o2}X_o \hat{\underset{\sim}{\gamma}} + A_{o2}FA_{o1}'(A_{o1}FA_{o1}'+V_1)^{-1}(\underset{\sim}{y}_{o1} - A_{o1}X_o\hat{\underset{\sim}{\gamma}})$$

$$= \underset{\sim}{a}_2'X_o \hat{\underset{\sim}{\gamma}} + \underset{\sim}{F}_{21}(F_{11}+U_{11})^{-1}(\hat{\underset{\sim}{\pi}}_{o1} - X_{o1} \hat{\underset{\sim}{\gamma}}) ;$$

where

$$\hat{\underset{\sim}{\pi}}_{o1} = (A'_{11}A_{11})^{-1}A'_{11}\underset{\sim}{y}_{o1} \; ; \; U_{11} = (A'_{11}V_1^{-1}A_{11})^{-1} \; ;$$

and $F$ and $X_o$ are partitioned accordingly as

$$F = \begin{bmatrix} F_{11} & \underset{\sim}{F}_{12} \\ \underset{\sim}{F}_{21} & F_{22} \end{bmatrix} \; ; \; X_o = \begin{bmatrix} X_{o1} \\ X_{o2} \end{bmatrix} .$$

It is not difficult to see that the increase in MSE due to uncertainty about $\underset{\sim}{\gamma}$ is

$$E(\underset{\sim}{\tilde{y}}_{o2} - \hat{y}_{o2})^2 = H_1 \{ var(\hat{\underset{\sim}{\gamma}}) \} H_1 \; ;$$

where

$$H_1 = [\underset{\sim}{a}'_{o2}X_o - \underset{\sim}{F}_{21}(F_{11}+U_{11})^{-1} X_{o1}] .$$

The predictor of $y_{o2}$ when all the parameters have to be estimated from the sample is given by

$$y^*_{o2} = \underset{\sim}{a}'_{o2}X_o \underset{\sim}{\hat{\gamma}} + \hat{\underset{\sim}{F}}_{21}(\hat{F}_{11}+\hat{U}_{11})^{-1}(\hat{\underset{\sim}{\pi}}_{o1} - X_{o1} \underset{\sim}{\hat{\gamma}}) .$$

Proceeding in the manner similar to what has been used in deriving (3.3.7) it follows that

$$E(y^*_{o2} - \underset{\sim}{\tilde{y}}_{o2})^2 \simeq N^{-1}[Tr(Q_{11}(F_{11}+U_{11}))][(F_{22}+\sigma_m^2) - \underset{\sim}{F}_{21}(F_{11}+U_{11})^{-1} \underset{\sim}{F}_{12}]$$

where

$$Q_{11} = (F_{11}+U_{11})^{-1} [ (F_{11}+U_{11})+X_{o1}\{ var(\hat{\underset{\sim}{\gamma}})\} X'_{o1}](F_{11}+U_{11})^{-1} .$$

So the approximate prediction mean square error is obtained as

$$E(y^*_{o2} - y_{o2})^2 \simeq \Omega_{2.1} + H_1\{ var(\hat{\underset{\sim}{\gamma}})\} H'_1$$

$$+ N^{-1} [ Tr(Q_{11}(F_{11}+U_{11}))] [ F_{22}+\sigma_m^2 - \underset{\sim}{F}_{21}(F_{11}+U_{11})^{-1}\underset{\sim}{F}_{12}] ,$$

$$(3.3.9)$$

MSE estimator of $\underset{\sim}{\pi}_i$ with MSE matrix given by

$$E\left[(\underset{\sim}{\tilde{\pi}}_i - \underset{\sim}{\pi}_i)(\underset{\sim}{\tilde{\pi}}_i - \underset{\sim}{\pi}_i)'\right] = U - U \, \Lambda^{-1} U.$$

When $\underset{\sim}{\gamma}$, $\Lambda$ and $\sigma_j^2$'s are unknown and are to be estimated from the data then it is natural to consider the estimator of $\underset{\sim}{\pi}_i$ as $\underset{\sim}{\tilde{\pi}}_i(c_1)$, given by

$$\underset{\sim}{\tilde{\pi}}_i(c_1) = \underset{\sim}{\hat{\pi}}_i - c_1 \, \hat{U} \, \hat{\Lambda}^{-1}(\underset{\sim}{\hat{\pi}}_i - X_i \, \underset{\sim}{\tilde{\gamma}}),$$

where constant $c_1$ is chosen in such a way that $E\left[\Sigma(\underset{\sim}{\tilde{\pi}}_i(c_1) - \underset{\sim}{\pi}_i)(\underset{\sim}{\tilde{\pi}}_i(c_1) - \underset{\sim}{\pi}_i)'\right]$ is minimized. It is very difficult to obtain this expected value. Instead, we consider the estimator of $\underset{\sim}{\pi}_i$ as $\underset{\sim}{\pi}_i^*(c)$, given by

$$\underset{\sim}{\pi}_i^*(c) = \underset{\sim}{\hat{\pi}}_i - c \, \hat{U} \, S_\lambda^{-1}(\underset{\sim}{\hat{\pi}}_i - X_i \, \underset{\sim}{\tilde{\gamma}}). \tag{3.4.1}$$

It is still difficult to obtain the expression for $E\left[\Sigma(\underset{\sim}{\pi}_i^*(c) - \underset{\sim}{\pi}_i)(\underset{\sim}{\pi}_i^*(c) - \underset{\sim}{\pi}_i)'\right]$. But a considerable amount of simplification is achieved if we assume that $S_\lambda$ is approximately distributed as $W(\Lambda; N-\bar{r}, q)$, where $\bar{r} = m^{-2} \sum_k \sum_\ell r_{k\ell}$. In the above $\bar{r}$ is taken as the average of $r_{k\ell}$'s. In the case considered by Reinsel (1984a) all the $r_{k\ell}$'s are the same and the distribution is exact. When $\bar{r}$ is not an integer, an integer nearest to it can be taken. Under this approximation it is not difficult to see that

$$E\left[\Sigma(\underset{\sim}{\pi}_i^*(c) - \underset{\sim}{\pi}_i)(\underset{\sim}{\pi}_i^*(c) - \underset{\sim}{\pi}_i)'\right] \simeq NU - \left[\frac{c^2}{N-\bar{r}-q-1} - 2c\right] U \, \Lambda^{-1} U.$$

The optimum value of c is $(N-\bar{r}-q-1)$, giving the minimum value

of $E\left[\Sigma(\pi_i^*(c)-\pi_i)(\pi_i^*(c)-\pi_i)'\right]$

$$= NU - (N-\bar{r}-q-1)\ U\ \Lambda^{-1}\ U$$

$$= NU - c\ U\Lambda^{-1}\ U \qquad (3.4.2)$$

If we do not introduce a constant multiplier in (3.4.1) but only substitute the estimates of the parameters then the estimator of $\pi_i$ will become

$$\hat{\pi}_i - (N-\bar{r})\hat{U}\ S_\lambda^{-1}\ (\hat{\pi}_i - X_i\ \tilde{\gamma})$$

$$= \pi_i^*\ (N-\bar{r}). \qquad (3.4.3)$$

3.5  MONTE CARLO STUDY

To compare the performance of the noniterative estimators with ML estimators a Monte Carlo simulation is performed. The performance of predictors based on these estimators and several others is also studied.

The data considered have two characteristics $(m = 2)$, one measured at five time points $(p_1=5)$ and the other at four time points $(p_2=4)$. The individual growth curves are linear for both the characteristics $(q_1=q_2=2)$, i.e. $\pi_1$ and $\pi_2$ are vectors of dimension 2. At the second stage $\pi_1$ has covariates as $x_1' = (1, z_1), z_1$ taking values $-1, 0$ and $1$ same number of times. For $\pi_2$ the covariate is $z_2$ taking values $1$ and $-2$ in the ratio $2:1$. The dimension of $\gamma$ is 6. The model of Section 3.2 with the following specifications is considered.

$$\underset{\sim}{\gamma}' = (50, 3, 11, 2, 3, 1.2);$$

$$X_i = \begin{bmatrix} 1 & z_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & z_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & z_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & z_2 \end{bmatrix} ; \quad A_1' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} ;$$

$$A_2' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} ; \quad A = \text{diag} (A_1, A_2).$$

The errors at stage two are chosen in such a way that the actual variance of $(\underset{\sim}{\pi}_i | \underset{\sim}{x}_i)$, $i = 1, 2$, explains a predetermined percent of the final variance of $\pi_i$. The errors at the first stage, i.e. $\sigma_1^2$ and $\sigma_2^2$ are chosen such that $(\underset{\sim}{y}_i | A_i)$, $i = 1, 2$, explains, on an average, a predetermined percent of the final variance of $\underset{\sim}{y}_i$.

The data are generated in a manner described above. For generating error vectors the subroutine GGNRM of ISML, which generates multivariate normal random vectors with $\underset{\sim}{0}$ mean and specified dispersion matrix, is used. We compare two types of estimators viz., the ML estimators (Section 3.2) and the non-iterative estimators (Section 3.4). To study the performance of the predictors based on estimates of the parameters obtained by these two methods, data for sixth time point of the first characteristic, i.e. for $A_{o2} = (1, 6, 0, 0)$ are generated. We consider the conditional prediction, i.e. prediction at $A_{o2}$ given the observed values at all the time points corresponding to the

design matrix A. The following five predictors are considered
for scrutiny.

(i) $\quad A_{o2}\,\hat{\underset{\sim}{\pi}}_i$ ;

(ii) $\quad A_{o2}\,\tilde{\underset{\sim}{\pi}}_i$ ;

(iii) $\quad A_{o2}\,\underset{\sim}{\pi}_i^*(c)$ ;

(iv) $\quad A_{o2}\,\underset{\sim}{\pi}_i^*(N-\bar{r})$ ;

(v) $\quad A_{o2}\,[\,\hat{\underset{\sim}{\pi}}_i - \hat{U}\,\hat{\Lambda}^{-1}(\hat{\underset{\sim}{\pi}}_i - X_i\,\hat{\underset{\sim}{\gamma}})\,]$ .

Predictor (i) is based on least square estimator of individual
$\underset{\sim}{\pi}_i$ and (ii) is the minimum MSE predictor obtained by putting
the known values of the parameters. Predictor (iii) is obtained
by using (3.4.1) whereas (iv) is obtained by using (3.4.3).
The predictor (v) is based on ML estimates of the parameters.

The prediction MSE of (i) and (ii) are given by
$(\sigma_1^2 + A_{o2}UA'_{o2})$ and $\Omega_{2.1}$, respectively. The prediction MSE of
(iii) and (iv) can be obtained by using (3.4.2), and that of (v)
by using (3.3.8). The estimated values of the prediction MSE
are obtained by substituting suitable estimates in place of the
parameters.

Four situations are considered for simulation study.
In the first two cases we consider a moderate sample size of 21
and in the other two cases a large sample size of 90 is
considered. For each sample size two situations are considered;
in one the percent of variability explained is 90 at both the
stages and in the other this level is 60 percent. For each of

these cases, 100 runs are carried out. These values are reported in Tables 3.1 - 3.3.

## 3.6 RESULTS OF MONTE CARLO STUDY

3.6.1 <u>Results on estimation of parameters</u> : It is evident from the Table 3.1 that the estimates of $\underset{\sim}{\gamma}$ are fairly satisfactory for both the methods. The componentwise mean square deviation (MSD) over 100 simulation runs for the noniterative method is smaller than that for the ML method in all the cases except for the third component of $\underset{\sim}{\gamma}$ when sample size is 90. However, overall they are not very different. For the same percent of variability explained, the larger sample size produces smaller componentwise MSD. For a given sample size, MSD decreases as the percent of explained variability increases. Both the results are on the expected lines.

Table 3.2 contains the estimates of F, $\sigma_1^2$ and $\sigma_2^2$. The estimates of $\sigma_1^2$ and $\sigma_2^2$ are the same for both the methods and so there is nothing to compare. The estimated values are close to the population values showing that the estimates are satisfactory.

It is difficult to compare the estimated value of F obtained by these two methods. For sample of size 21, all the components of the mean estimated F obtained by the noniterative method are nearer (in the sense of absolute difference) to the true F than those by ML method. However, for the samples of size 90 the position is not clear. The estimate of $F_{11}$, the 2x2

Table 3.1:Estimate of $\gamma_2$ by Two Methods viz., the Maximum Likelihood (ML) and the Noniterative(NI), and their Mean Square Deviations (MSD)

| Popula-tion $\gamma_2$ | Sample Size 21 | | | | | | | | Sample Size 90 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 90% Variability Explained | | | | 60% Variability Explained | | | | 90% Variability Explained | | | | 60% Variability Explained | | | |
| | ML | NI | MSD (ML) | MSD (NI) | ML | NI | MSD (ML) | MSD (NI) | ML | NI | MSD (ML) | MSD (NI) | ML | NI | MSD (ML) | MSD (NI) |
| 50.000 | 50.003 | 49.969 | 0.213 | 0.202 | 50.005 | 49.921 | 1.355 | 1.290 | 49.979 | 49.966 | 0.036 | 0.035 | 49.944 | 49.908 | 0.274 | 0.264 |
| 3.000 | 2.995 | 2.970 | 0.677 | 0.608 | 2.882 | 2.923 | 4.284 | 3.864 | 2.967 | 2.966 | 0.118 | 0.110 | 2.907 | 2.902 | 0.893 | 0.832 |
| 11.000 | 10.992 | 10.991 | 0.027 | 0.023 | 10.982 | 10.978 | 0.172 | 0.145 | 11.008 | 11.011 | 0.004 | 0.004 | 11.021 | 11.029 | 0.030 | 0.030 |
| 2.000 | 2.006 | 2.005 | 0.084 | 0.080 | 2.015 | 2.012 | 0.526 | 0.501 | 2.009 | 2.009 | 0.015 | 0.014 | 2.023 | 2.023 | 0.103 | 0.102 |
| 3.000 | 2.934 | 2.886 | 0.407 | 0.372 | 2.813 | 2.678 | 3.214 | 2.962 | 3.012 | 3.009 | 0.089 | 0.085 | 3.031 | 3.074 | 0.647 | 0.621 |
| 1.200 | 1.255 | 1.264 | 0.052 | 0.051 | 1.353 | 1.379 | 0.414 | 0.400 | 1.198 | 1.199 | 0.013 | 0.013 | 1.195 | 1.198 | 0.094 | 0.094 |

Table 3.2:Estimates of F, $\sigma_1^2$ and $\sigma_2^2$ by Two Methods viz., the Maximum Likelihood (ML) and the Noniterative(NI)

| | 90% Variability Explained | | | | | 60% Variability Explained | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Population Values | Sample Size 21 ML | NI | Sample Size 90 ML | NI | Population Values | Sample Size 21 ML | NI | Sample Size 90 ML | NI |
| F(1,1) | 0.333 | 0.058 | 0.374 | 0.274 | 0.361 | 1.999 | 0.237 | 2.250 | 1.544 | 2.211 |
| F(2,1) | 0.000 | 0.074 | 0.005 | 0.024 | 0.002 | 0.000 | 0.481 | -0.028 | 0.171 | 0.001 |
| F(3,1) | 0.000 | 0.044 | 0.029 | -0.010 | -0.010 | 0.000 | 0.280 | 0.187 | -0.060 | -0.060 |
| F(4,1) | 0.000 | -0.001 | 0.005 | -0.007 | -0.007 | 0.000 | -0.019 | 0.022 | -0.051 | -0.052 |
| F(2,2) | 0.148 | 0.105 | 0.143 | 0.138 | 0.149 | 0.889 | 0.617 | 0.854 | 0.810 | 0.894 |
| F(2,3) | 0.000 | -0.025 | -0.019 | -0.007 | -0.007 | 0.000 | -0.158 | -0.119 | -0.045 | -0.047 |
| F(2,4) | 0.000 | 0.002 | 0.000 | 0.005 | 0.005 | 0.000 | 0.020 | 0.002 | 0.032 | 0.032 |
| F(3,3) | 0.333 | 0.220 | 0.328 | 0.330 | 0.356 | 2.000 | 1.075 | 1.926 | 2.003 | 2.189 |
| F(3,4) | 0.000 | 0.039 | 0.007 | -0.009 | -0.017 | 0.000 | 0.326 | 0.060 | -0.068 | 0.124 |
| F(4,4) | 0.053 | 0.036 | 0.052 | 0.056 | 0.060 | 0.320 | 0.184 | 0.304 | 0.342 | 0.368 |
| $\sigma_1^2$ | 3.112 | 3.118 | 3.118 | | | 19.947 | 19.988 | 19.998 | | |
| $\sigma_2^2$ | 1.565 | 1.550 | 1.550 | | | 12.852 | 12.726 | 12.726 | | |
| $\sigma_1^2$ | 3.372 | | | 3.372 | 3.372 | 26.114 | | | 26.119 | 26.119 |
| $\sigma_2^2$ | 1.399 | | | 1.370 | 1.370 | 10.352 | | | 10.133 | 10.133 |

Table 3.3:Mean Square Deviation(MSD) and Mean Square Error(MSE) for Predictors (i) to (v)

|  | Sample Size 21 | | | | | | Sample Size 90 | | | | | |
| | 90% Variability Explained | | | 60% Variability Explained | | | 90% Variability Explained | | | 60% Variability Explained | | |
| Predictors | MSD | MSE | Ratio | MSD | MSE | Ratio | MSD | MSE | Ratio | MSD | MSE | Ratio |
| (i) | 6.511 | 6.548 | 1.006 | 41.739 | 41.975 | 1.006 | 7.209 | 7.082 | 0.984 | 55.829 | 54.849 | 0.982 |
| (ii) | 4.778 | 4.595 | 0.962 | 30.434 | 29.262 | 0.962 | 5.016 | 4.940 | 0.985 | 37.829 | 37.262 | 0.985 |
| (iii) | 8.501 | 4.101 | 0.482 | 40.035 | 26.020 | 0.650 | 5.399 | 4.850 | 0.898 | 37.296 | 36.556 | 0.930 |
| (iv) | 6.490 | 4.754 | 0.733 | 40.613 | 30.275 | 0.745 | 7.118 | 4.977 | 0.699 | 55.066 | 37.599 | 0.683 |
| (v) | 10.914 | 4.315 | 0.395 | 43.294 | 27.411 | 0.633 | 5.516 | 4.896 | 0.888 | 39.510 | 36.931 | 0.935 |

dispersion matrix of the regression parameters for the first characteristic, obtained by the noniterative method is nearer to its population value than the one obtained by the ML method. But for $F_{22}$, the 2x2 dispersion matrix corresponding to the regression parameters for the second characteristic, the ML method gives the estimate that is nearer to its population value than by the noniterative method. Nothing of this sort is evident for $F_{12}$. We may say that for small sizes the non-iterative method behaves more satisfactorily than the ML method. For large sample size, however, the ML method improves the estimate of F considerably. This is as expected because the ML method generally does better for large sample sizes.

3.6.2  <u>Results on predictors</u> : Results on prediction are summarized in Table 3.3. The comparison of predictors is based on MSD. In all the four cases the minimum MSE predictor with known parameters (ii) performs the best (i.e. has the smallest MSD). The predictor based on the noniterative estimates without introducing a constant multiplier (iv) is better than the predictor based on individual $\hat{\pi}_i$ (i). The introduction of constant multiplier c (iii) gives better results than (iv), as one will expect, in three cases. But for the sample of size 21 and 90 percent variability explained, we see that the introduction of c increases the MSD value. This is the case where even (i) performs better than (iii). This observation does not fit in to the general pattern. For sample of size 21 the predictor based

on the ML estimates has the highest MSD. However, for sample
of size 90 this predictor does quite well as one will expect
because ML estimates are satisfactory for large sample size.
In this case, its MSD is smaller than for predictor (iv) but
slightly larger than (iii).

Looking at the Table 3.3 we observe that larger sample
size produces the ratio of the MSE and MSD which is nearer to
unity than for smaller sample size for predictors (ii), (iii)
and (v). Irrespective of sample size this ratio is close to
one for predictors (i) and (ii). For the predictor (v) this
ratio is far from one for small sample size but is near one for
large sample size. This is as expected since the expression
(3.3.8) gives MSE in the asymptotic case. The behaviour of the
predictor (iii) is similar to that of (v).

## 3.7 ILLUSTRATIVE EXAMPLE

In this section we explain the method by working out
an example. The data are taken from a survey conducted by
Indian Council of Agricultural Research on preharvest forecast
of yield of sugarcane in Meerut district in India in the year
1976-77 (see Experiment 4). The data consist of measurements
of cane heights (in meters) and diameters (in centimeters) of
randomly selected plants from 33 plots of 25 squares meters
each. Five records are available on height and diameter from
each sampled plant. Final yield of canes, giving the total
weight of sugarcane, and number of shoots for each plot are also
available.

A second degree polynomial in time appears to give
an adequate fit for the height. A second degree polynomial
for the diameter is also found adequate. However, the diameters
do not show significant correlation with yield and are not
considered in great detail for the present analysis. For
sugarcane yield, number of shoots per plot is considered as
a covariate. For height no covariates are considered. For
the $i$-th plot $\underset{\sim}{y}_{i1}$ is a $5 \times 1$ vector corresponding to height at
5 time points, $y_{i2}$ gives final yield of the plot and $n_i$ gives
the number of shoots in the $i$-th plot. In terms of the above
notation

$$\underset{\sim}{y}_{i1} = A_1 \, \underset{\sim}{\pi}_{i1} + \underset{\sim}{e}_{i1} \; ; \quad y_{i2} = \pi_{i2} + e_{i2} \; ,$$

where $k$-th row of $A_1$ is $(1, k, k^2)$;

$$E(\underset{\sim}{\pi}'_{i1}) = (\gamma_1, \gamma_2, \gamma_3) \; ; \quad E(\pi_{i2}) = \gamma_4 + \gamma_5 \, n_i \; ;$$

$$\underset{\sim}{\pi}_i = (\underset{\sim}{\pi}'_{i1}, \pi_{i2})' \; ; \quad \text{var}(\underset{\sim}{e}_{i1}) = \sigma_1^2 I \; ; \quad \text{var}(e_{i2}) = \sigma_2^2 \; ; \quad \text{var}(\underset{\sim}{\pi}_i) = F .$$

The comprehensive model is given by

$$\underset{\sim}{y}_i = A X_i \underset{\sim}{\gamma} + A \, \underset{\sim}{\eta}_i + \underset{\sim}{e}_i \; ,$$

where

$$A = \begin{array}{c} 5 \\ \\ 1 \end{array} \begin{bmatrix} A_1 & \underset{\sim}{0} \\ \\ \underset{\sim}{0}' & 1 \end{bmatrix} \; ; \quad X_i = \begin{array}{c} 3 \\ \\ 1 \end{array} \begin{bmatrix} I & \underset{\sim}{0} \\ \\ \underset{\sim}{0}' & 1, n_i \end{bmatrix} .$$
$$\qquad\quad 3 \quad 1 \qquad\qquad\qquad 3 \quad 2$$

The maximum likelihood estimates of $\underset{\sim}{\gamma}$, $\Lambda$ and $\sigma_1^2$ are obtained by iteration and convergence does not create any special problem. In this case $\sigma_2^2$ is not estimable explicitly. The estimates are

$$\hat{\underset{\sim}{\gamma}}' = (0.38, 0.49, -0.04, 17.10, 0.25) \; ;$$

$$\text{var}(\hat{\underset{\sim}{\gamma}}) = 10^{-2} \times \begin{bmatrix} 0.581 & & & & \\ -0.339 & 0.266 & & & \\ 0.046 & -0.337 & 0.005 & & \\ 1.758 & 0.330 & 0.055 & 1216.884 & \\ 0.000 & 0.000 & 0.000 & -9.308 & 0.080 \end{bmatrix} \; ;$$

$$\hat{\Lambda} = \begin{bmatrix} 0.192 & & & \\ -0.112 & 0.088 & & \\ 0.015 & -0.012 & 0.002 & \\ 0.581 & -0.109 & 0.018 & 42.099 \end{bmatrix} \; ; \quad \hat{\sigma}_1^2 = 0.007$$

No test is made to check the adequacy of the above model. A visual inspection of simple covariance matrix of heights and that obtained from the above fitted model do not show much discrepancy. Inspection of $\hat{\Lambda}$ suggests that second and third components of $\underset{\sim}{\pi}_{i1}$ can be considered as fixed parameters.

The above growth curve (G.C.) Model was used for conditional prediction. The yield $y_{2i}$ of the i-th plot is predicted using heights upto time t $(t = 1, \ldots, 5)$ and is denoted by $y_{2ti}$. The value of MSD, given by $N^{-1}\Sigma(y_{2i} - y_{2ti}^*)^2$ where summation extends over all values of i, has been calculated. We also evaluate corresponding value using expression (3.3.9)

and compare it with the MSD value (see Table 3.4).

For comparison purposes another simple predictor of $Y_{2i}$, obtained by regressing height at each stage and number of shoots, is calculated. Such predictors based on multiple regression (M.R.) method are in current use. They, however, do not include any information on previous stage because of complications due to collinearity. Growth curve approach for prediction has many other advantages over M.R. approach. For the latter one has to develop different prediction formulae for different stages. Moreover, equations developed for one set of points are not suitable for predictions when data are not recorded exactly at those time points. However, the prediction from G.C. approach is free from such limitations.

For these data there does not appear to be any systematic decrease in the prediction error as data on more time points become available. This is also true with M.R. method. This is perhaps due to small size of N in comparison to the number of parameters estimated. When all the auxiliary information due to heights is ignored then the MSD is 42.10 which shows that heights do not reduce the MSD considerably in this case.

When diameters are also taken into account then the MSD for G.C. method is 30.71 while that for M.R. is 37.88, corresponding to the fifth stage. M.R. method takes into account the height and diameter data of single stage and number of shoots. When data on all the stages are taken into account,

Table 3.4: Mean Square Error of Predictors for Multiple

Regression(MR) and Growth Curve(GC) Methods.

| Stage | MSD (MR) | MSD (GC) | MSE (GC) |
|---|---|---|---|
| 3 | 38.63 | 37.33 | 43.29 |
| 4 | 36.55 | 35.70 | 43.16 |
| 5 | 30.64 | 37.25 | 42.91 |

the MSD of M.R. method is considerably reduced to 16.62. This, however, cannot be relied upon as the number of parameters is 12 from 33 observations.

G.C. method may be preferable over M.R. method for the reasons given above, though for the present data there is not much to choose between them on the basis of MSD.

## 3.8 DISCUSSION AND SUMMARY

In this chapter we discuss a generalised RCR model which allows different design matrix and different dimension of regression parameters for each characteristic. The random regression coefficients for each characteristic may also depend upon different set of auxiliary variables. This generalises the results of Reinsel (1984a) in many directions. Such models have potential applications in many situations. A possible application in preharvest forecasting has been considered in great details. Working of the proposed method is illustrated by using real data in Section 3.7.

The maximum likelihood procedure is adopted for estimation of parameters. The estimates are obtained by iteration. Optimum properties of ML estimators have also been discussed. Some simple noniterative and unbiased estimators, which may provide starting values for iteration, are given in Section 3.4. The results of simulation study indicate that for estimation of $\gamma$, the noniterative and ML estimators are close to the true values and both the methods are satisfactory. For small sample size the

noniterative method provides more satisfactory estimate of F than the ML method. For large sample size, however, the ML method produces much improved estimate of F as can be expected.

Problems of conditional prediction are discussed in detail in Section 3.3. An asymptotic expression for MSE of prediction (see equation (3.3.8)) is derived. On the basis of simulation study carried out in Section 3.5 we find that this approximation is satisfactory for large sample size. In simulation study the performance of various predictors, including the empirical Bayes type predictor and predictor based on ML estimates, are also compared. For small sample size the performance of the empirical Bayes type predictor is superior to that of the one based on ML estimates. For large sample size both are close competitors.

Reinsel (1982,1984a) has considered this model with the same design matrix for all individuals and same covariates for all the characteristics. He has obtained least square estimator of $\underset{\sim}{\gamma}$ which is also ML estimator and is noniterative in nature. The noniterative estimators considered in Section 3.4 are similar to the least square estimators of Reinsel. From the above discussion it appears that this method has a clear advantage over ML method for the parameters considered here, apart from the simplicity in computation, atleast for moderate sample size.

The studies on simulation experiments are not exhaustive. Moreover, the simulation results are based on a very small number

of experiments. The study is taken only to confirm the theoretical results and in most of the cases considered here, the simulation results agree well with what is expected on the theoretical grounds.

# APPENDIX 3.1

## Derivation of expression (3.3.7)

$$\tilde{\underset{\sim}{y}}_{02} - \underset{\sim}{y}^{*}_{02} = A_{02}\{\hat{U}_1(\hat{F}+\hat{U}_1)^{-1} - U_1(F+U_1)^{-1}\}\ (\underset{\sim}{\pi}_0 - X_0\,\underset{\sim}{\hat{\gamma}}).$$

Now we use the following approximation,

$$\hat{U}_1(\hat{F}+\hat{U}_1)^{-1} - U_1(F+U_1)^{-1} = \hat{U}_1(F+U_1)^{-1} - U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)(F+U_1)^{-1}$$

$$+ \{U_1(F+U_1)^{-1} - \hat{U}_1(\hat{F}+\hat{U}_1)^{-1}\}\{(\hat{F}+\hat{U}_1)-(F+U_1)\}(F+U_1)^{-1}$$

$$\simeq \{\hat{U}_1 - U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)\}(F+U_1)^{-1}$$

$$E\{(\tilde{\underset{\sim}{y}}_{02} - \underset{\sim}{y}^{*}_{02})(\tilde{\underset{\sim}{y}}_{02} - \underset{\sim}{y}^{*}_{02})'\}$$

$$\simeq A_{02}\left[E\{\hat{U}_1 - U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)\}\,Q\,\{\hat{U}_1 - U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)\}'\right]A'_{02}$$

$$= A_{02}\left[E\{\hat{U}_1 Q\hat{U}_1\} - E\{\hat{U}_1 Q(\hat{F}+\hat{U}_1)(F+U_1)^{-1}U_1\} - E\{U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)Q\hat{U}_1\}\right.$$

$$\left. + E\{U_1(F+U_1)^{-1}(\hat{F}+\hat{U}_1)Q(\hat{F}+\hat{U}_1)(F+U_1)^{-1}U_1\}\right]A'_{02}$$

where

$$Q = (F+U_1)^{-1}\left[(F+U_1)+X_0\{\mathrm{var}(\underset{\sim}{\hat{\gamma}})\}X'_0\right](F+U_1)^{-1} = (Q_{ij}).$$

Next,

$$E(\hat{U}_1 Q\hat{U}_1) = U_1 Q U_1 + P \tag{i}$$

where

$$P = \mathrm{diag}(g_1 P_1,\ldots,g_m P_m)\ ;\ g_j = 2\{N(p_j - q_j)\}^{-1}\sigma_j^4\ ,\ j=1,2,\ldots,m;$$

$$P_j = (A'_{1j}A_{1j})^{-1}Q_{jj}(A'_{1j}A_{1j})^{-1}.$$

Here $A_{1j}$ is the design matrix corresponding to the observed part of $j$-th characteristic in $\underset{\sim}{y}_{o1}$.

$$E\left[(\hat{F}+\hat{U}_1)Q(\hat{F}+\hat{U}_1)\right] = E\left[\hat{\Lambda}Q\hat{\Lambda}+(\hat{U}_1-\hat{U})Q\hat{\Lambda}+\hat{\Lambda}Q(\hat{U}_1-\hat{U})+(\hat{U}_1-\hat{U})Q(\hat{U}_1-\hat{U})\right]$$

$$= \Lambda Q\Lambda + N^{-1}\left[\Lambda Q\Lambda + Tr(\Lambda Q)\Lambda\right] + (U_1-U)Q\Lambda + \Lambda Q(U_1-U)$$

$$+ (U_1-U)Q(U_1-U) + D$$

$$= (F+U_1)Q(F+U_1)+N^{-1}\left[\Lambda Q\Lambda +Tr(\Lambda Q)\Lambda\right] + D \qquad \text{(ii)}$$

where

$$D = \left[(A_1'A_1)^{-1} - (A'A)^{-1}\right]Q\left[(A_1'A_1)^{-1} - (A'A)^{-1}\right].$$

$$E\left[\hat{U}_1Q(\hat{F}+\hat{U}_1)\right] = E\left[\hat{U}_1Q\hat{\Lambda} + \hat{U}_1Q(\hat{U}_1-\hat{U})\right]$$

$$= U_1Q\Lambda + U_1Q(U_1-U)+P-T$$

$$= U_1Q(F+U_1)+P-T \qquad \text{(iii)}$$

where

$$T = (A_1'A_1)^{-1}Q\left[(A_1'A_1)^{-1}-(A'A)^{-1}\right].$$

Using (i), (ii) and (iii) we obtain

$$E\left[(\underset{\sim}{\tilde{y}}_{o2}-\underset{\sim}{y}_{o2}^*)(\underset{\sim}{\tilde{y}}_{o2}-\underset{\sim}{y}_{o2}^*)'\right]$$

$$\simeq A_{o2}\left[N^{-1}\{U_1(F+U_1)^{-1}\Lambda Q\Lambda(F+U_1)^{-1}U_1+Tr\{\Lambda Q\}U_1(F+U_1)^{-1}\Lambda(F+U_1)^{-1}U_1\}\right.$$

$$+ P-U_1(F+U_1)^{-1}P-P(F+U_1)^{-1}U_1+U_1(F+U_1)^{-1}T+T(F+U_1)^{-1}U_1$$

$$+ \left. U_1(F+U_1)^{-1}D(F+U_1)^{-1}U_1\right]A_{o2}'.$$

CHAPTER    IV

# A TWO-STAGE AUTOREGRESSIVE MODEL FOR REPEATED MEASUREMENTS

## 4.1   INTRODUCTION

Most of the economic time series analysis is confined to asymptotic results based on a single large series of observations. However, in the study of biological and physical processes over time, it is more likely that data on replicated series over a relatively small interval of time may be available. It is of interest to estimate the parameters assuming that they remain same over units. A vast literature on growth curve analysis, wherein it is assumed that the parameters do not change over time, is available; for recent reviews see Ware (1983) and Khatri (1985). However, in some situations it may be more appropriate to consider that the underlying parameters change over time in some regular fashion. Anderson (1978) has considered a multivariate autoregressive process for repeated series and obtained the ML estimators of the parameters. He has also developed procedures to test the nature of autoregressive parameter matrix.

From a practical point of view, there is a need to generalise Anderson's (1978) single stage model to a two-stage model. To have some insight into situations where such generalisations may be required, consider the problem of

predicting yield of perennial trees. The yield of a tree may depend on some auxiliary variables such as age, weather condition and may be expressed by a regression relationship for each year. Moreover, the regression coefficients for each tree may change from year to year in a regular fashion due to process of aging, which may be represented by a multivariate autoregressive process. In this chapter we have considered such a two-stage model for repeated measurements.

The two-stage model that we discuss here is similar to the well known Kalman filter model, e.g. see Kalman (1960), Duncan and Horn (1972), Sallas and Harville (1981), Meinhold and Singpurwalla (1983) and West et al. (1985). In Kalman filter models there is only one series of observations and the transition matrix is assumed to be known. This may be a valid assumption for processes governed by physical laws but for biological processes the transition matrix may not be known. Availability of a number of series enables one to estimate the transition matrix as in Anderson (1978). Other related work with random coefficient time varying parameters in the context of time series is reported by Rosenberg (1973 a, b). Swamy and Mehta (1977), Swamy and Tinsley (1981) and references therein.

In Section 4.2 we propose a two-stage model. An EM algorithm for the calculation of maximum likelihood (ML) estimates of the model parameters is derived in Section 4.3. A noniterative method of estimation, which may provide initial

values for ML procedure, is suggested in Section 4.4. Section 4.5 deals with the problem of prediction. Section 4.6 describes a Monte Carlo simulation study. The results of simulation study are discussed in Section 4.7. An illustrative example is presented in Section 4.8. The discussion and summary are given in Section 4.9.

## 4.2 MODEL

Consider a two-stage model in which at the first-stage an observable $p_{\alpha t} \times 1$ random vector $\underset{\sim}{y}_{\alpha t}$ for the $\alpha$th individual at time $t$ is expressed by a regression relationship

$$\underset{\sim}{y}_{\alpha t} = X_{\alpha t} \underset{\sim}{\beta}_{\alpha t} + \underset{\sim}{e}_{\alpha t} , \quad \alpha = 1,2,\ldots,N; \quad t = 1,2,\ldots,T;$$

$$(4.2.1)$$

where $X_{\alpha t}$ is a $p_{\alpha t} \times q$ design matrix; $\underset{\sim}{\beta}_{\alpha t}$ is a $q \times 1$ vector of random regression parameters, $\underset{\sim}{e}_{\alpha t}$ is a $p_{\alpha t} \times 1$ error vector. We assume that

$$E(\underset{\sim}{e}_{\alpha t}) = \underset{\sim}{0} , \quad E(\underset{\sim}{e}_{\alpha t} \underset{\sim}{e}'_{\alpha' t'}) = \sigma^2 \delta_{\alpha \alpha'} \delta_{t t'} I ,$$

where $\delta$ is the Kronecker delta and $\sigma^2$ is a positive scalar; $\underset{\sim}{\beta}_{\alpha t}$ and $\underset{\sim}{e}_{\alpha' t'}$ are uncorrelated for all $\alpha$, $\alpha'$ and for all $t$, $t'$.

At the second stage we assume that $\underset{\sim}{\beta}_{\alpha t}$ follows a first-order Markov process with zero mean and may be expressed as

$$\underset{\sim}{\beta}_{\alpha t} = G \underset{\sim}{\beta}_{\alpha t-1} + \underset{\sim}{\eta}_{\alpha t} , \quad \alpha = 1,2,\ldots,N, \quad t = 2,3,\ldots,T; \quad (4.2.2)$$

where $G$ is a $q \times q$ matrix and $\{\underset{\sim}{\eta}_{\alpha t}\}$ is a sequence of $q \times 1$ independent (unobservable) random vectors with

$$E(\underset{\sim}{\eta}_{\alpha t}) = \underset{\sim}{0} \; ; \; E(\underset{\sim}{\eta}_{\alpha t} \, \underset{\sim}{\eta}'_{\alpha' t'}) = \delta_{\alpha \alpha'} \delta_{tt'} \, V \, ,$$

where $V$ is a symmetric positive definite matrix.

We also assume that $\underset{\sim}{\beta}_{\alpha 1}$ is distributed normally as $N(0, \underset{\sim}{R}_1)$. When characteristic roots of $G$ are less than one in absolute value then the process is stationary.

In the above formulation it has been assumed that mean of $\underset{\sim}{y}_{\alpha t}$'s is zero. In practice the mean may not be zero and it may have to be estimated from the data. Some modifications are required in the estimation procedure when mean is nonzero and are given in Section 4.3.2.

Anderson (1978) has considered only a single stage model corresponding to relationship (4.2.2) but making $G$ and $V$ depending upon time as $G_t$ and $V_t$. The two-stage extension of this model is similar to Kalman's filter models. In Kalman's filter model $G_t$ and $V_t$ are assumed to be known which may be a valid assumption for the processes governed by physical laws but in biological processes $G_t$ and $V_t$ may have to be estimated from the data. In this study we have considered the case when $G_t$ and $V_t$ remain same over time as well as over replications. Availability of number of replications enables one to estimate $G$ and $V$ even for a relatively short series, as in the case of Anderson (1978).

## 4.3 ESTIMATION

Define a $p_\alpha (= \sum\limits_{t=1}^{T} p_{\alpha t})$ vector $\underset{\sim}{y}_\alpha$ and a $T \cdot q$ vector $\underset{\sim}{\beta}_\alpha$ as

$$\underset{\sim}{y}_\alpha = (\underset{\sim}{y}'_{\alpha 1}, \underset{\sim}{y}'_{\alpha 2}, \ldots, \underset{\sim}{y}'_{\alpha T})' \; ; \; \underset{\sim}{\beta}_\alpha = (\underset{\sim}{\beta}'_{\alpha 1}, \underset{\sim}{\beta}'_{\alpha 2}, \ldots, \underset{\sim}{\beta}'_{\alpha T})'.$$

Assuming errors $\underset{\sim}{e}_{\alpha t}$ and $\eta_{\alpha t}$ to be normally distributed, the conditional probability density of $\underset{\sim}{y}_\alpha$, given $\underset{\sim}{\beta}_\alpha$, can be written as

$$(\underset{\sim}{y}_\alpha | \underset{\sim}{\beta}_\alpha) \sim N(X_\alpha \underset{\sim}{\beta}_\alpha, \sigma^2 I), \tag{4.3.1}$$

and the marginal density of $\underset{\sim}{\beta}_\alpha$ as

$$\underset{\sim}{\beta}_\alpha \sim N(0, \Lambda), \tag{4.3.2}$$

where

$$X_\alpha = \text{diag}(X_{\alpha 1}, X_{\alpha 2}, \ldots, X_{\alpha T}) \; ; \; \Lambda = U V_* U' \; ;$$

where $U$ is a lower triangular matrix with elements

$$U = (U_{ij}) = (G^{i-j}), \; i > j \text{ and } V_* = \text{diag}(R_1, V, \ldots, V).$$

Combining (4.3.1) and (4.3.2) and integrating over $\underset{\sim}{\beta}_\alpha$'s, the likelihood of parameters, given $\underset{\sim}{y}_\alpha$'s, $\alpha = 1, 2, \ldots, N$; after some manipulations, can be written as

$$L = \prod_{\alpha=1}^{N} f(\underset{\sim}{y}_\alpha | R_1, V, \sigma^2, G)$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}Tp} \cdot |R_1|^{-\frac{1}{2}N} \cdot |V|^{-\frac{1}{2}(T-1)N} \cdot (\prod_{\alpha=1}^{N} |C_\alpha|^{-\frac{1}{2}})$$

$$\times \exp\{-\frac{1}{2}\sum_{\alpha=1}^{N}(\sigma^{-2}\underset{\sim}{y}'_\alpha \underset{\sim}{y}_\alpha - \underset{\sim}{\varphi}'_\alpha C_\alpha^{-1} \underset{\sim}{\varphi}_\alpha)\}, \tag{4.3.3}$$

where

$$p = \sum_{\alpha=1}^{N} p_\alpha ; \; C_\alpha = \Lambda^{-1} + \sigma^{-2} \text{diag}(X'_{\alpha 1} X_{\alpha 1}, X'_{\alpha 2} X_{\alpha 2}, \ldots, X'_{\alpha T} X_{\alpha T}) ;$$

$$\underset{\sim}{\varphi}_\alpha = \sigma^{-2} X'_\alpha \underset{\sim}{y}_\alpha.$$

The likelihood is a complicated function of parameters and the ML solutions by differentiation are not easy to obtain. An alternative method, which does not require the use of likelihood function as such, is via EM algorithm (see Dempster et al., 1977) and is described in the next subsection.

4.3.1 <u>ML estimates by EM algorithm</u> : To describe the steps of the EM algorithm for the model of Section 4.2, rewrite the equations (4.3.1) and (4.3.2) as

$$
\begin{bmatrix} \underset{\sim}{y}_\alpha \\ \underset{\sim}{\beta}_\alpha \end{bmatrix} \sim N \left( \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{bmatrix} , \begin{bmatrix} X_\alpha \Lambda X'_\alpha + \sigma^2 I & X_\alpha \bar{\Lambda} \\ \Lambda X'_\alpha & \Lambda \end{bmatrix} \right) , \alpha = 1, 2, \ldots, N.
$$

$$(4.3.4)$$

After $\underset{\sim}{y}_\alpha$ has been observed inference about $\underset{\sim}{\beta}_\alpha$, when $\Lambda$ and $\sigma^2$ are given, is characterised by the conditional distribution of $\underset{\sim}{\beta}_\alpha$ given $\underset{\sim}{y}_\alpha$, i.e.

$$
\underset{\sim}{\beta}_\alpha | \underset{\sim}{y}_\alpha \sim N(\underset{\sim}{\beta}^*_\alpha , D_\alpha)
$$

where

$$
\underset{\sim}{\beta}^*_\alpha = \Lambda X'_\alpha (X_\alpha \Lambda X'_\alpha + \sigma^2 I)^{-1} \underset{\sim}{y}_\alpha , \tag{4.3.5}
$$

$$
D_\alpha = \Lambda - \Lambda X'_\alpha (X_\alpha \Lambda X'_\alpha + \sigma^2 I)^{-1} X_\alpha \Lambda . \tag{4.3.6}
$$

If $\underset{\sim}{\beta}_\alpha$'s were observed we would then estimate $G, R_1, V$ and $\sigma^2$ as follows :

$$
\underset{\sim}{g}^* = vec(G^*)
$$

$$
= \{ I \otimes ( \sum_{t=2}^{T} \sum_{\alpha} \underset{\sim}{\beta}_{\alpha t-1} \underset{\sim}{\beta}'_{\alpha t-1} )^{-1} \} \{ \sum_{t=2}^{T} \sum_{\alpha} (\underset{\sim}{\beta}_{\alpha t} \otimes \underset{\sim}{\beta}'_{\alpha t-1}) \} , \tag{4.3.7}
$$

$$N(T-1)V^* = \sum_{t=2}^{T} \sum_{\alpha} (\underset{\sim}{\beta}_{\alpha t} - G\underset{\sim}{\beta}_{\alpha t-1})(\underset{\sim}{\beta}_{\alpha t} - G\underset{\sim}{\beta}_{\alpha t-1})' \, , \qquad (4.3.8)$$

$$NR_1^* = \sum_{\alpha} \underset{\sim}{\beta}_{\alpha 1} \underset{\sim}{\beta}'_{\alpha 1} \, , \qquad (4.3.9)$$

$$Tp\sigma^{2*} = \sum_{t=1}^{T} \sum_{\alpha} (\underset{\sim}{y}_{\alpha t} - X_{\alpha t} \underset{\sim}{\beta}_{\alpha t})'(\underset{\sim}{y}_{\alpha t} - X_{\alpha t} \underset{\sim}{\beta}_{\alpha t}) \, , \qquad (4.3.10)$$

where $\text{vec}(A)$ is a column vector formed by stacking the rows of $A$, and $\otimes$ is a Kronecker product of matrices. These estimates cannot be formed because $\underset{\sim}{\beta}_{\alpha}$'s are not known.

Let $\underset{\sim}{\theta}$ be a vector of unknown parameters. The EM algorithm forms $\underset{\sim}{\theta}^{(k+1)}$ as a solution of $(4.3.7) - (4.3.10)$ using the expected values of the moment matrices of $\underset{\sim}{\beta}_{\alpha t}$'s, conditional on $\underset{\sim}{\theta}^{(k)}$ and the observed data. It is not difficult to see that

$$E(\underset{\sim}{\beta}_{\alpha t}\underset{\sim}{\beta}'_{\alpha t} | \underset{\sim}{y}_{\alpha}, \underset{\sim}{\theta}^{(k)}) = \underset{\sim}{\beta}_{\alpha t}^{(k)*} \underset{\sim}{\beta}_{\alpha t}^{(k)*'} + D_{\alpha,tt}^{(k)} \, ;$$

$$E(\underset{\sim}{\beta}_{\alpha t}\underset{\sim}{\beta}'_{\alpha t-1} | \underset{\sim}{y}_{\alpha}, \underset{\sim}{\theta}^{(k)}) = \underset{\sim}{\beta}_{\alpha t}^{(k)*} \underset{\sim}{\beta}_{\alpha t-1}^{(k)*'} + D_{\alpha,tt-1}^{(k)} \, ;$$

$$E\left[(\underset{\sim}{y}_{\alpha t} - X_{\alpha t}\underset{\sim}{\beta}_{\alpha t})'(\underset{\sim}{y}_{\alpha t} - X_{\alpha t}\underset{\sim}{\beta}_{\alpha t}) | \underset{\sim}{y}_{\alpha}, \underset{\sim}{\theta}^{(k)}\right]$$

$$= (\underset{\sim}{y}_{\alpha t} - X_{\alpha t}\underset{\sim}{\beta}_{\alpha t}^{(k)*})'(\underset{\sim}{y}_{\alpha t} - X_{\alpha t}\underset{\sim}{\beta}_{\alpha t}^{(k)}) + \text{Tr}(X'_{\alpha t}X_{\alpha t}D_{\alpha,tt}^{(k)}) \, ;$$

where $D_{\alpha} = (D_{\alpha,ij})$ and each $D_{\alpha,ij}$ is a $q \times q$ matrix and $\underset{\sim}{\beta}_{\alpha t}$ is a suitable vector component of $\underset{\sim}{\beta}_{\alpha}$.

Now all the terms necessary for forming the conditional expected value and the EM algorithm are completely defined. The estimation step is given by $(4.3.5)$ and $(4.3.6)$ and the conditional

expected values of the terms on the right hand side of (4.3.7)
(4.3.10) provide the maximization step.

4.3.2 <u>Modifications when mean is not zero</u> : We suggest
modifications to be made when the mean of $\underset{\sim}{y}_{\alpha t}$ is different
from zero. Consider a general mixed model of the form

$$\underset{\sim}{y}_{\alpha t} = z_{\alpha t}\underset{\sim}{\mu} + X_{\alpha t}\underset{\sim}{\beta}_{\alpha t} + \underset{\sim}{e}_{\alpha t} ,$$

$$\underset{\sim}{\beta}_{\alpha t} = G\underset{\sim}{\beta}_{\alpha t-1} + \underset{\sim}{\eta}_{\alpha t} .$$

In such a setup one has to estimate additional parameters $\underset{\sim}{\mu}$ from
the data. After $\underset{\sim}{y}_{\alpha}$ has been observed inference about $\underset{\sim}{\beta}_{\alpha}$ , when
$\Lambda$ , $\sigma^2$ and $\underset{\sim}{\mu}$ are given, is characterised by the conditional
distribution of $\underset{\sim}{\beta}_{\alpha}$ given $\underset{\sim}{y}_{\alpha}$ , i.e.

$$(\underset{\sim}{\beta}_{\alpha} | \underset{\sim}{y}_{\alpha}) \sim N(\overset{\sim}{\underset{\sim}{\beta}}_{\alpha} , D_{\alpha})$$

where

$$\overset{\sim}{\underset{\sim}{\beta}}_{\alpha} = \Lambda X'_{\alpha}(X_{\alpha} \Lambda X'_{\alpha} + \sigma^2 I)^{-1} (\underset{\sim}{y}_{\alpha} - Z_{\alpha} \underset{\sim}{\mu}) , \qquad (4.3.11)$$

$$Z_{\alpha} = (Z'_{\alpha 1}, \dots, Z'_{\alpha T})' .$$

If $\underset{\sim}{\beta}_{\alpha}$'s were known $G, R_1, V, \sigma^2$ and $\underset{\sim}{\mu}$ are estimated as the follow-

$$\overset{\sim}{\underset{\sim}{g}} = vec(\overset{\sim}{G})$$
$$= \{ I \otimes ( \sum_{t=2}^{T} \sum_{\alpha} \beta_{\alpha t-1} \beta'_{\alpha t-1} )^{-1} \} . \{ \sum_{t=2}^{T} \sum_{\alpha} (\underset{\sim}{\beta}_{\alpha t} \otimes \underset{\sim}{\beta}'_{\alpha t-1}) \} \qquad (4.3.12)$$

$$N(T-1)\overset{\sim}{V} = \sum_{t=2}^{T} \sum_{\alpha} (\underset{\sim}{\beta}_{\alpha t} - G\underset{\sim}{\beta}_{\alpha t-1})(\underset{\sim}{\beta}_{\alpha t} - G\underset{\sim}{\beta}_{\alpha t-1})' , \qquad (4.3.13)$$

$$N\overset{\sim}{R}_1 = \sum_{\alpha} \underset{\sim}{\beta}_{\alpha 1} \underset{\sim}{\beta}'_{\alpha 1} , \qquad (4.3.14)$$

$$Tp\tilde{\sigma}^2 = \sum_{t=1}^{T} \sum_{\alpha} (y_{\alpha t} - z_{\alpha t}\tilde{\mu} - x_{\alpha t}\beta_{\alpha t})'(y_{\alpha t} - z_{\alpha t}\tilde{\mu} - x_{\alpha t}\beta_{\alpha t}), \quad (4.3.15)$$

$$N\tilde{\mu} = \sum_{\alpha} \{ (z_\alpha' z_\alpha)^{-1} z_\alpha' (y_\alpha - x_\alpha\beta_\alpha) \}. \quad (4.3.16)$$

The conditional expected values of the terms on the right hand side of equations (4.3.12) – (4.3.16) can be obtained without much difficulty.

4.3.3 Comments on the properties of the estimators : As we are dealing with multivariate normal distribution, usual assumptions for the ML estimators to have optimum properties are satisfied. Thus the estimators are consistent for large N. However, for fixed N, and T tending to infinite, one may have to impose condition of statinarity and knowledge of $R_1$ for proving the consistency of the estimators. As for as the information matrix is concerned one may have to obtain it by numerical differentiation as analytical expressions are very complicated.

In the present problem the possibility of process converging to a local maximum cannot be ruled out. Under these circumstancesthe choice of starting value can be very crucial. If one starts with initial estimators which are consistent, then under certain usual assumptions of the maximum likelihood this process is bound to converge to estimators which are consistent and asymptotically efficient (Lehmann, 1983, Chapter 6). In the next section we describe a method for obtaining consistent estimators which can be taken as initial values for the algorithm.

## 4.4 A NONITERATIVE METHOD OF ESTIMATION

In the above section we have considered the maximum likelihood estimation procedure which is iterative. To initiate the iteration initial guess values of the parameters are required. In this section we describe a method of obtaining noniterative estimators, which may provide good guess values of the parameters, when all $X_{\alpha t}$'s are equal to, say, X and $p_{\alpha t}$ $(= p_0$, say$) > q$. We also discuss their properties. In this case $X_\alpha$ reduces to $X_* = \text{diag}(X, X, \ldots, X)$ and correspondingly $C_\alpha$ to $C$.

The likelihood of parameters given $y_\alpha$'s is given by (4.3.3). After some matrix manipulation it is not difficult to see that (4.3.3) can equivalently be written as

$$L = (2\pi\sigma^2)^{-\frac{1}{2}Tp} \cdot \exp\left[-\tfrac{1}{2}\sigma^{-2} \Sigma Q(\hat{\beta}_\alpha)\right]$$

$$\times |R_1|^{-\frac{1}{2}N} \cdot |V|^{-\frac{1}{2}(T-1)N} \cdot |C|^{-\frac{1}{2}N} \cdot$$

$$\times \exp\left[-\tfrac{1}{2}\sigma^{-2} \Sigma \hat{\beta}'_\alpha \{(X'_* X_*)^{-1} + \sigma^2 \Lambda\}^{-1} \hat{\beta}_\alpha\right]$$

$$= L_1 \times L_2$$

where

$$Q(\hat{\beta}_\alpha) = (y_\alpha - X_* \hat{\beta}_\alpha)'(y_\alpha - X_* \hat{\beta}_\alpha),$$

$$\hat{\beta}_\alpha = (X'_* X_*)^{-1} X'_* y_\alpha.$$

From this it is clear that $T_1 = \Sigma Q(\hat{\beta}_\alpha)$ and $T_2 = \Sigma \hat{\beta}_\alpha \hat{\beta}'_\alpha$ are sufficient statistics. Given $\beta_\alpha$, the distribution of

$\Sigma \, Q(\hat{\underset{\sim}{\beta}}_\alpha)/\sigma^2$ has a $\chi^2$ distribution on $p-Nq$ degrees of freedom and is distributed independently of $\hat{\underset{\sim}{\beta}}_\alpha$. From the nature of the likelihood it is evident that, given $\sigma^2$, parameters in $\Lambda$ can be estimated by $T_2$ only. An unbiased estimator of $\sigma^2$ is obtained as

$$\hat{\sigma}^2 = \Sigma \, Q(\hat{\underset{\sim}{\beta}}_\alpha) \div p-Nq.$$

For structured matrix $\Lambda$, which is a function of $G, R_1$ and $V$, the estimators of parameters cannot be obtained in a closed form. Even the differentiation with respect to the parameters $G, R_1$ and $V$ is not easy and do not result in equations that may be easy to solve. However, considering $L_2$ only it can be seen that

$$L_2 = |R_1|^{-\frac{1}{2}N} \, |V|^{-\frac{1}{2}(T-1)N} \, \cdot |C|^{-\frac{1}{2}N}$$

$$\cdot \exp\left[-\tfrac{1}{2}\sigma^2 \, \mathrm{Tr}\{(I+\sigma^2 P'(X'_* X_*)^{-1}P)(\Sigma \, P'\hat{\underset{\sim}{\beta}}_\alpha \hat{\underset{\sim}{\beta}}'_\alpha P)\}\right]$$

where $P$ is such that $\Lambda^{-1} = PP'$. It can be seen that

$$\hat{\underset{\sim}{\beta}}'_\alpha P = \left[\hat{\underset{\sim}{\beta}}'_{\alpha 1}R_1^{-\frac{1}{2}}, (\hat{\underset{\sim}{\beta}}_{\alpha 2}-G\hat{\underset{\sim}{\beta}}_{\alpha 1})' V^{-\frac{1}{2}}, \ldots, (\hat{\underset{\sim}{\beta}}_{\alpha T}-G\hat{\underset{\sim}{\beta}}_{\alpha T-1})' V^{-\frac{1}{2}}\right].$$

In cases where all the latent roots of $P(X'_* X_*)^{-1}P$ are less than unity in absolute value, we may expand $[I+\sigma^2 P'(X'_* X_*)^{-1}P]^{-1}$ in the powers of $P'(X'_* X_*)^{-1}P$. The dominating term in the exponent term of $L_2$ is

$$\Sigma \, \hat{\underset{\sim}{\beta}}'_\alpha PP'\hat{\underset{\sim}{\beta}}_\alpha = \Sigma \, \hat{\underset{\sim}{\beta}}'_{\alpha 1}R_1^{-1}\hat{\underset{\sim}{\beta}}_{\alpha 1} + \overset{T}{\underset{t=2}{\Sigma}} \, \underset{\alpha}{\Sigma} \, (\hat{\underset{\sim}{\beta}}_{\alpha t}-G\hat{\underset{\sim}{\beta}}_{\alpha t-1})' V^{-1}(\hat{\underset{\sim}{\beta}}_{\alpha t}-G\hat{\underset{\sim}{\beta}}_{\alpha t-1}).$$

which is the same as in the expression of likelihood in Anderson (1978) with $\underset{\sim}{y}_\alpha$'s replaced by $\hat{\underset{\sim}{\beta}}_\alpha$'s. The estimators of $R_1$, G and V can be obtained by expressions similar to (3.7), (3.12) and (3.13) of Anderson (1978), respectively, i.e.

$$R_{1c} = C_1(0) \; ;$$

$$G_c = C(1) \cdot C^*(0)^{-1} \; ; \tag{4.4.1}$$

$$N(T-1)V_c = \sum_{t=2}^{T} \Sigma \, (\hat{\underset{\sim}{\beta}}_{\alpha t} - G_c \hat{\underset{\sim}{\beta}}_{\alpha t-1})(\hat{\underset{\sim}{\beta}}_{\alpha t} - G_c \hat{\underset{\sim}{\beta}}_{\alpha t-1})' \; ; \tag{4.4.2}$$

where

$$(T-1)C(j) = \sum_{t=2}^{T} C_t(j) \; ;$$

$$(T-1)C^*(j) = \sum_{t=2}^{T} C_{t-1}(0) \; ;$$

$$NC_t(j) = \Sigma \, \hat{\underset{\sim}{\beta}}_{\alpha t} \, \hat{\underset{\sim}{\beta}}_{\alpha t-j}' \; .$$

From the definition of $\hat{\underset{\sim}{\beta}}_\alpha$ it follows that

$$E(\hat{\underset{\sim}{\beta}}_{\alpha t}) = \underset{\sim}{0}; \; var(\hat{\underset{\sim}{\beta}}_{\alpha t}) = R_t + \sigma^2(X'X)^{-1}; \; cov(\hat{\underset{\sim}{\beta}}_{\alpha t}, \hat{\underset{\sim}{\beta}}_{\alpha t-1}) = GR_{t-1}.$$

Now it can be seen that

$$E[C_1(0)] = R_1 + \sigma^2(X'X)^{-1} \; ;$$

$$E[C(1)] = G(R_1 + R_2 + \cdots + R_{T-1}) \; ;$$

$$E[C^*(0)] = [R_1 + R_2 + \cdots + R_{T-1} + (T-1)\sigma^2(X'X)^{-1}] \; .$$

The above expressions suggest that $R_1$ and $\sum_{t=1}^{T-1} R_t$ can be estimated unbiasedly by $M_1$ and $M_2$, respectively, where

$$M_1 = C_1(0) - \hat{\sigma}^2 (X'X)^{-1}; \quad M_2 = C^*(0) - (T-1) \hat{\sigma}^2 (X'X)^{-1}.$$

However, in some cases both $M_1$ and $M_2$ may fail to be positive definite. A better estimator of G can thus be obtained by replacing $C^*(0)$ by $M_2$ in (4.4.1). The estimator is given by

$$\hat{G} = C(1) \, M_2^{-1}.$$

By using law of large numbers it is easy to show that for fixed T, $C(1)$ and $M_2$ tend to their expected values i.e. as $N \to \infty$

$$C(1) \xrightarrow{\text{p}} G(R_1 + \dots + R_{T-1});$$

$$M_2 \xrightarrow{\text{p}} (R_1 + \dots + R_{T-1}).$$

Thus $\hat{G}$ gives a consistent estimator of G. However, for finite $N, M_2$ may turn out to be nearly singular in some cases, and in such cases it may be advisable to modify it so as to make it invertible.

## 4.5   PREDICTION

Suppose $\underset{\sim}{y}_\alpha$ has been observed upto time T and one wishes to predict the values of $\underset{\sim}{y}_\alpha$ at time T+h. If G were known then the minimum mean square error predictor is just the conditional mean, given by

$$\underset{\sim}{y}^*_{\alpha T+h} = X_{T+h} \, G^h \, \underset{\sim}{\beta}^*_{\alpha T} ,$$

where $\underset{\sim}{\beta}^*_{\alpha T}$ is the vector of last q components of $\underset{\sim}{\beta}^*_\alpha = E(\underset{\sim}{\beta}_\alpha | \underset{\sim}{y}_\alpha)$, given in (4.3.5).

The prediction variance is

$$\text{var}(\underset{\sim}{y}^{*}_{\alpha T+h}) = X_{T+h} \left[ \sum_{i=0}^{h-1} G^{i} V G^{\prime i} + G^{h} R_{T} G^{\prime h} \right] X^{\prime}_{T+h} + \sigma^{2} I, \quad (4.5.1)$$

Since G is not known its estimate is substituted to give the predicted value of $\underset{\sim}{y}_{\alpha T+h}$ as

$$\underset{\sim}{\tilde{y}}_{\alpha T+h} = X_{T+h} G^{*h} \underset{\sim}{\beta}^{*}_{\alpha T}.$$

It is not easy to evaluate the prediction variance of $\underset{\sim}{\tilde{y}}_{\alpha T+h}$. The prediction variance evaluated by substituting the estimated values for parameter values in (4.5.1) will be an underestimate because it does not take into account the variability due to using estimated value of G.

## 4.6  MONTE CARLO SIMULATION STUDY

To compare the performance of the simple noniterative estimators and the ML estimators, a Monte Carlo simulation study is performed. The data are generated in the following manner.

The data vectors of dimension 4 (i.e. $p_{\alpha t} = 4$ for all $\alpha$ and t) are generated for T=4. The regression relationship at the first stage is considered to be linear (i.e. q = 2). The design matrices $X_{\alpha t}$'s are taken to be the same for all $\alpha$ and t as

$$X^{\prime}_{\alpha t} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \end{bmatrix}.$$

The $\underset{\sim}{\beta}_{\alpha 1}$'s are considered to have been generated from a bivariate normal distribution with mean $\underset{\sim}{0}$ and dispersion matrix $R_{1}$, given by

$$R_1 = \begin{bmatrix} 0.40 & 0.00 \\ 0.00 & 0.40 \end{bmatrix}.$$

Two cases are considered for second stage model viz.,
the stationary and the nonstationary. In the stationary case
we have $R_1 = R_2 = \ldots = R_t = R$ (say). This imposes the condition that
the characteristic roots of G be less than 1 in absolute value
and V satisfy the conditions that $V = R - GRG'$. The autoregressive
matrix G is taken to be

$$G = \begin{bmatrix} 0.60 & 0.00 \\ 0.00 & 0.70 \end{bmatrix},$$

giving V as

$$V = \begin{bmatrix} 0.256 & 0.000 \\ 0.000 & 0.204 \end{bmatrix},$$

which is symmetric and positive definite. For the nonstationary
case G is taken as

$$G = \begin{bmatrix} 1.20 & 0.00 \\ 0.00 & 1.32 \end{bmatrix}.$$

Two different values of V viz.,

$$V = \begin{bmatrix} 0.256 & 0.000 \\ 0.000 & 0.204 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0.16 & 0.00 \\ 0.00 & 0.16 \end{bmatrix},$$

are chosen to enable comparison over V. Moreover, the value

$$V = \begin{bmatrix} 0.256 & 0.000 \\ 0.000 & 0.204 \end{bmatrix},$$

is the same as for the stationary case so that the comparison over the values of G can be made.

The amount of variability at the first stage $(\sigma^2)$ is chosen in such a way that, on the average, $(\underset{\sim}{y}_{\alpha t} | X_{\alpha t})$ explains a predetermined percent of the total error. For the stationary case four situations are considered; a moderate sample size of 30 and a large sample size of 95, each combining with two levels (60% and 90%) of variability explained at the first stage. For the nonstationary case we restrict the study to 90% variability explained at the first stage. Two values of V and two sample sizes of 30 and 95 give four combinations of parameters.

For generating random normal vectors the subroutine GGNRM of IMSL, which generates multivariate normal random vectors with $\underset{\sim}{0}$ mean and specified dispersion matrix, is used. We compare two types of estimates viz., the ML estimates via EM algorithm (Section 4.3) and the noniterative estimates (Section 4.4), on the basis of outcome of 100 simulation runs. These are reported in Tables 4.1 - 4.2.

## 4.7 RESULTS OF SIMULATION STUDY

The results of Monte Carlo simulation over 100 runs are summarised in Tables 4.1 - 4.2. Perusal at these tables reveals that for estimation of parameters $G, R_1, V$ and $\sigma^2$, both the procedures are satisfactory for stationary as well as nonstationary cases. There was no need for modification of $M_2$ as it was always well behaved for the purposes of inversion. It is observed that

Table 4.1:Estimates of Parameters G,R₁,V and $\sigma^2$ by Two Methods viz.,(1) the Noniterative(NI) and (11)the Maximum

likelihood (ML) and their Mean Square Deviations(MSD) for Stationary Case

| | Population Value | Sample Size 30 60% Variability Explained Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Sample Size 30 90% Variability Explained Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Sample Size 95 60% Variability Explained Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Sample Size 95 90% Variability Explained Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G(1,1) | 0.60 | 0.61 | 0.58 | 6.23 | 20.4 | 0.59 | 0.59 | 0.97 | 0.89 | 0.61 | 0.61 | 1.59 | 1.24 | 0.60 | 0.60 | 0.29 | 0.29 |
| G(1,2) | 0.00 | -0.02 | 0.05 | 2.28 | 32.5 | -0.01 | -0.01 | 0.75 | 0.73 | 0.00 | 0.00 | 0.62 | 0.51 | 0.00 | 0.00 | 0.24 | 0.24 |
| G(2,1) | 0.00 | -0.01 | -0.02 | 1.84 | 9.54 | 0.00 | 0.00 | 0.70 | 0.69 | -0.01 | -0.02 | 0.50 | 0.39 | 0.00 | 0.00 | 0.19 | 0.18 |
| G(2,2) | 0.70 | 0.71 | 0.66 | 1.06 | 18.4 | 0.69 | 0.69 | 0.57 | 0.57 | 0.70 | 0.70 | 0.29 | 0.26 | 0.70 | 0.70 | 0.16 | 0.16 |
| R₁(1,1) | 0.40 | 0.36 | 0.37 | 4.25 | 3.84 | 0.38 | 0.38 | 1.42 | 1.40 | 0.38 | 0.38 | 1.46 | 1.37 | 0.39 | 0.39 | 0.51 | 0.53 |
| R₁(2,1) | 0.00 | 0.00 | -0.01 | 0.98 | 1.14 | 0.00 | 0.00 | 0.48 | 0.48 | 0.00 | 0.00 | 0.32 | 0.26 | 0.00 | 0.00 | 0.16 | 0.15 |
| R₁(2,2) | 0.40 | 0.39 | 0.39 | 1.65 | 1.82 | 0.39 | 0.39 | 1.21 | 1.21 | 0.40 | 0.40 | 0.40 | 0.38 | 0.40 | 0.40 | 0.31 | 0.31 |
| V(1,1) | 0.26 | 0.25 | 0.26 | 1.94 | 1.62 | 0.26 | 0.26 | 0.23 | 0.23 | 0.24 | 0.25 | 0.60 | 0.58 | 0.25 | 0.25 | 0.08 | 0.08 |
| V(2,1) | 0.00 | 0.00 | 0.00 | 0.33 | 0.26 | 0.00 | 0.00 | 0.08 | 0.07 | 0.00 | 0.00 | 0.11 | 0.10 | 0.00 | 0.00 | 0.03 | 0.03 |
| V(2,2) | 0.20 | 0.19 | 0.19 | 0.26 | 0.29 | 0.20 | 0.20 | 0.12 | 0.12 | 0.20 | 0.20 | 0.08 | 0.07 | 0.20 | 0.20 | 0.04 | 0.04 |
| $\sigma^2$ | 1.60 | 1.59 | 1.62 | 1.97 | 3.52 | | | | | 1.60 | 1.60 | 0.65 | 0.68 | | | | |
| $\sigma^2$ | 0.27 | | | | | 0.27 | 0.27 | 0.06 | 0.06 | | | | | 0.27 | 0.27 | 0.02 | 0.02 |

Table 4.2:Estimates of $G$, $R_1$, $V$ and $\sigma^2$ by Two Methods viz.,(i) the Noniterative(NI) and (ii) the Maximum

Likelihood (ML) for Nonstationary Case when 90 % Variability is Explained by the Model

| | Popul ation Value | Sample Size 30 Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Sample Size 95 Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Popul ation Value | Sample Size 30 Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 | Sample Size 95 Estimate NI | ML | MSD (NI) x100 | MSD (ML) x100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G(1,1) | 1.20 | 1.21 | 1.18 | 1.23 | 0.65 | 1.21 | 1.20 | 0.38 | 0.36 | 1.20 | 1.21 | 1.15 | 1.26 | 0.86 | 1.21 | 1.19 | 0.38 | 0.48 |
| G(1,2) | 0.00 | -0.01 | 0.00 | 0.64 | 0.61 | -0.01 | 0.00 | 0.18 | 0.17 | 0.00 | -0.01 | -0.01 | 0.55 | 0.61 | -0.01 | 0.00 | 0.14 | 0.17 |
| G(2,1) | 0.00 | -0.01 | -0.01 | 0.56 | 0.74 | 0.00 | 0.01 | 0.12 | 0.12 | 0.00 | -0.01 | 0.00 | 0.54 | 0.73 | -0.01 | -0.01 | 0.12 | 0.23 |
| G(2,2) | 1.32 | 1.32 | 1.31 | 0.34 | 0.34 | 1.32 | 1.32 | 0.09 | 0.09 | 1.32 | 1.32 | 1.31 | 0.29 | 0.28 | 1.32 | 1.32 | 0.08 | 0.08 |
| R₁(1,1) | 0.40 | 0.37 | 0.38 | 2.72 | 2.11 | 0.38 | 0.39 | 1.01 | 0.94 | 0.40 | 0.37 | 0.39 | 2.55 | 1.98 | 0.38 | 0.39 | 0.95 | 0.94 |
| R₁(2,1) | 0.00 | 0.00 | -0.01 | 0.75 | 0.65 | 0.00 | 0.00 | 0.24 | 0.18 | 0.00 | 0.00 | 0.00 | 0.71 | 0.62 | 0.00 | 0.00 | 0.23 | 0.18 |
| R₁(2,2) | 0.40 | 0.39 | 0.40 | 1.45 | 1.39 | 0.40 | 0.40 | 0.36 | 0.33 | 0.40 | 0.35 | 0.40 | 1.42 | 1.36 | 0.40 | 0.40 | 0.35 | 0.33 |
| V(1,1) | 0.26 | 0.26 | 0.30 | 2.24 | 1.31 | 0.25 | 0.25 | 0.71 | 0.50 | 0.16 | 0.17 | 0.26 | 1.71 | 2.09 | 0.15 | 0.18 | 0.52 | 0.66 |
| V(1,2) | 0.00 | 0.00 | 0.00 | 0.40 | 0.35 | 0.00 | 0.01 | 0.12 | 0.09 | 0.00 | 0.00 | 0.00 | 0.29 | 0.29 | 0.00 | 0.07 | 0.08 | 0.11 |
| V(2,2) | 0.20 | 0.19 | 0.20 | 0.29 | 0.21 | 0.20 | 0.20 | 0.09 | 0.08 | 0.16 | 0.15 | 0.16 | 0.21 | 0.13 | 0.16 | 0.16 | 0.07 | 0.05 |
| $\sigma^2$ | 1.02 | 1.02 | 1.01 | 0.81 | 0.83 | 1.02 | 1.02 | 0.27 | 0.38 | 0.95 | 0.94 | 0.93 | 0.69 | 0.60 | 0.94 | 0.95 | 0.23 | 0.48 |

on the basis of MSD there is no indication that one method is uniformly better than the other. The noniterative procedure is computationary very simple and provides very good estimates. Infact, noniterative estimators give more satisfactory results than the ML procedure particularly when sample size is small. These results are very much on expected lines. The EM algorithm locates the region of convergence very quickly but later on the convergence of this procedure is rather slow.

For stationary case it is observed that larger percent of variability explained by the model produces smaller MSD values. Same is the case with sample size i.e. larger sample size produces smaller MSD. Both the observations follow expected patterns.

For nonstationary case we do not study the effect of changing the percent of variability explained by the model. Instead, we study the variation due to change in V. No systematic change in MSD is observed by making V smaller. In this case also the larger sample size yields smaller MSD as one will expect.

## 4.8 AN EXAMPLE

To get some insight into the performance of $\tilde{\underset{\sim}{y}}_{\alpha T+h}$ as a predictor of $\underset{\sim}{y}_{\alpha T+h}$, we illustrate the method suggested in this chapter by a small example. The data reported by Rao (1981), which are originally due to Williams and Izenman (1981) (see Experiment 2), are used. The data are not adequate for the

present model but have been used for illustration purposes only. The data consist of weights of 13 mice measured at intervals of 3 days over 21 days beginning from the third day. Only a single measurement at each time is made giving $p_{\alpha t} = 1$ for all $\alpha$ and $t$. We consider a mixed model of the form

$$Y_{\alpha t} = \mu_0 + (2t-7)\mu_1 + \beta_{o\alpha t} + (2t-7)\beta_{1\alpha t} + e_{\alpha t} \quad , \quad \alpha = 1, \ldots, 13; \ t = 1, 2, \ldots, 6;$$

which may be written as

$$Y_{\alpha t} = (1, 2t-7)\underset{\sim}{\mu} + (1, 2t-7)\underset{\sim}{\beta}_{\alpha t} + e_{\alpha t}.$$

$\underset{\sim}{\beta}_{\alpha t}$'s are assumed to be related as

$$\underset{\sim}{\beta}_{\alpha t} = G\underset{\sim}{\beta}_{\alpha t-1} + \underset{\sim}{\eta}_{\alpha t}.$$

Parameters are estimated from the data upto 18 days. These values are used to predict the weight of each mouse on 21st day. The predicted weights are compared with the corresponding observed weight of each mouse.

The maximum likelihood estimates are obtained as follows:

$$\underset{\sim}{\mu} = (0.5793, \ 0.0654)' \ ; \ \overset{\sim}{\sigma}^2 = 0.5481 \times 10^{-4} \ ;$$

$$\overset{\sim}{G} = \begin{bmatrix} 1.2779 & -1.8779 \\ 0.1160 & 0.5088 \end{bmatrix} \ ; \ \overset{\sim}{R}_1 = 10^{-1} \times \begin{bmatrix} 0.2671 & 0.0674 \\ 0.0674 & 0.0176 \end{bmatrix} \ ;$$

$$\overset{\sim}{V} = 10^{-2} \times \begin{bmatrix} 0.1009 & 0.0217 \\ 0.0217 & 0.0049 \end{bmatrix} \ .$$

The EM algorithm locates the region of convergence rather quickly but later on the rate of convergence is very slow.

To get some idea how good the above dispersion structure fits the data, we have performed the likelihood ratio test. When there is no restriction on the dispersion matrix, the maximum likelihood estimators are given by

$$\hat{\mu} = (Z'S^{-1}Z)^{-1}Z'S^{-1}\bar{y} \; ; \; \widehat{var}(\underset{\sim}{y}) = [S+N(\bar{y}-Z\hat{\mu})(\bar{y}-Z\hat{\mu})']/N,$$

where S is the 6x6 corrected sum of squares and product matrix and $\bar{y}$ is the 6x1 sample mean vector. Number of parameters fitted in the unconstrained case is 23 and corresponding maximum value of the loglikelihood is 138.9376. For the above structured model 13 parameters are fitted giving the maximum value of the loglikelihood as 131.5795. This gives likelihood ratio test statistic

$$-2 \log \Lambda = 12.7162 ,$$

on 10 degrees of freedom, showing that it is not significant. Hence, it appears that this model gives a reasonable fit. The number of observations is too small and it is difficult to put much faith on the sensitivity of this test in the present case.

The estimated values are used to predict the weights of mice on 21st day (h=1). These values are then compared with the known observed weights on 21st day. The sum of squares of deviations, given by $\Sigma(y_{iT+h}-\tilde{y}_{iT+h})^2$, where $y_{iT+h}$ and $\tilde{y}_{iT+h}$ are observed and predicted weights of the ith mouse at (T+h)th time point, respectively, is 0.0381. This value is

considerably smaller than the corresponding value obtained by using best linear unbiased predictor, James-Stein regression predictor, simple regression predictor, empirial Bayes predictor. These values are reported by Rao (1981) and the minimum value is 0. 0336 for empirical Bayes predictor. The value reported in Rao (1981) is 0.0951 which is in some error due to omission of some factor (Personal communication with C.R. Rao).

The prediction variance obtained from (4.5.1) comes out to be 0.0335 which is considerably larger than the empirical mean square error value which is 0.0029. This inflated value of prediction variance may be due to the fact that many parameters have been estimated from a small data and hence the estimates may not be precise.

## 4.9   DISCUSSION AND SUMMARY

In this chapter, a two-stage model is considered in Section 4.2 wherein the first stage is the usual regression model and at the second stage the regression parameters of the first stage are assumed to follow autoregressive process. These models have possible applications in prediction in perennial crops and other biological processes. The resulting model is similar to the well known Kalman filter model. In the latter model only a single series is observed and consequently a number of parameters cannot be estimated and are assumed to be known. In the present work it is assumed that replicated series are available enabling us to estimate all the parameters.

For estimation of the parameters, the ML procedure via EM algorithm is adopted in Section 4.3. Some simple alternate estimators, which are consistent, are considered in Section 4.4. Performances of these two methods are studied on the basis of Monte Carlo simulations. It appears that the proposed alternative estimators perform very well when sample size is small. Even for large sample size their performance is satisfactory. Moreover, computationally they are much simpler in comparison to ML estimates.

The proposed model has been used for prediction purposes on a real set of data and results indicate that the model may have some potential use. However, a lot more additional work needs to be done before this model could be accepted as a potential alternative to the existing methods for the analysis of the longitudinal data. For example, it is very necessary to obtain the variances of estimators. A derivation for the MSE of prediction of predictors based on estimated values of the parameters or atleast some approximation thereof should be available. Some test for goodness of fit of the model to the data should also be available.

The number of simulation runs for all the cases is 100 which may not be sufficient for making assertive comments. The study is not exhaustive either. Our aim is merely to see the behaviour of various estimators and in most of the cases the observed nature of these estimators follow the lines one would expect from theoretical considerations.

CHAPTER V

ESTIMATION AND PREDICTION IN MIXED MODELS WITH
AUTOCORRELATED ERRORS

## 5.1 INTRODUCTION

A general family of mixed linear models proposed by
Harville (1976), which includes growth and repeated measurement
models, has been extensively studied in literature, for example
see Harville (1976,1977), Dempster et al. (1981), Laird and
Ware (1982) and Kackar and Harville (1984). In most of the
work related to growth and repeated measurement models it is
assumed that the errors are distributed as multivariate normal
with mean $\underset{\sim}{0}$ and dispersion matrix $\sigma^2 I$ or $\sigma^2 V$ with a known matrix
V. The assumption that the dispersion matrix of errors is $\sigma^2 I$
may, in some cases, be too restrictive. If V is a general
unknown matrix then it may not be possible to estimate V because
all the parameters may not be identifiable. It is of
considerable interest to consider a case when the dispersion
matrix of errors is of autoregressive nature in addition to
some coefficients being random. Some researchers have considered
such models, for example see Swamy (1971) and Mansour et al.
(1985).

A mixed linear model with autocorrelated errors is
proposed in Section 5.2 which may be more appropriate in
longitudinal studies. In Section 5.3 we obtain the maximum

likelihood (ML) estimators of the parameters which are iterative
in nature. The problem of prediction and derivation of its
mean square error (MSE) are considered in Section 5.4. Some
simple estimators, which provide good guess values to initiate
the iteration, are obtained in Section 5.5. Properties of
these estimators are also discussed. The description of Monte
carlo simulation is given in Section 5.6. The results of
simulation study are summarized in Section 5.7. An illustrative
example is given in Section 5.8. The work done in this chapter
is briefly reviewed in Section 5.9.

## 5.2 MODEL

Let us consider a mixed effect regression model given by

$$y_i = X\mu + Z_1\beta_i + e_i \ , \ i = 1,2,\ldots,N \ ; \qquad (5.2.1)$$

where $y_i$ is a px1 vector of observable random variables; X and
$Z_1$ are known design matrices of order pxm and pxq, respectively;
$\mu$ is a mx1 vector of unobservable fixed effects, $\beta_i$ is a qx1
vector of unobservable individual random effects with $E(\beta_i) = 0$
and $var(\beta_i) = F$ and $e_i$ is a px1 vector of errors.

The above mixed model includes growth and repeated
measurement models. For example, the random coefficient regression
model considered by Rao (1965,1967,1975) can be treated as
particular cases of this model. The situation considered by
Fearn (1977), where the dimensions of fixed and random effects
may be different, falls within the purview of the model
represented by (5.2.1). It is possible to consider a more

general version of this model where X and $Z_1$ differ from unit to unit. This type of generalised model is considered by Laird and Ware (1982) and Ware and Berkey (1984). Swamy (1971), Fearn (1975) and Rao (1975) consider this type of generalisation for $Z_1=X$ and X being different for different units. In the present work this generalisation has not been considered as it introduces lot of complexities in the autoregressive nature of dispersion matrix for $\underset{\sim}{e}$'s.

We consider a case where components of $\underset{\sim}{y}_i$ consist of observations taken at equal interval of time which may give rise to errors following a first order autoregressive process. For errors we assume that $\underset{\sim}{e}_i \sim N(0,\sigma^2 V(\rho))$ where

$$V(\rho) = (v_{ij}(\rho)) = ((1-\rho^2)^{-1} \rho^{|i-j|}).\qquad (5.2.2)$$

Combining (5.2.1) and (5.2.2), the model can be written as

$$\underset{\sim}{y}_i = X\underset{\sim}{\mu} + Z\underset{\sim}{b}_i$$

where

$$Z = (Z_1:I) \; ; \; \underset{\sim}{b}_i = (\underset{\sim}{\beta}_i':\underset{\sim}{e}_i')'.$$

It can be seen that

$$E(\underset{\sim}{y}_i) = X\underset{\sim}{\mu}; \; var(\underset{\sim}{y}_i) = Z D(\underset{\sim}{\theta})Z' = \Omega \; (say),$$

where $\underset{\sim}{\theta}$ is a vector of unknown parameters (i.e. $F, \sigma^2$ and $\rho$) and

$$D(\underset{\sim}{\theta}) = diag(F,\sigma^2 V(\rho)).$$

Swamy (1971, Section 4.4) has considered a model with $Z_1=X$ and $var(\underset{\sim}{e}_i) = \sigma^2 V(\rho_i)$. He has obtained consistent estimators

of components of $\underset{\sim}{\theta}$. His estimator of $\rho_i$ is satisfactory if large number of observations (i.e. p large) are available on each unit. However, in most longitudinal studies number of observations on each unit may hardly exceed eight or ten, but it is possible to choose N considerably large (Anderson, 1978). In many cases it may be realistic to assume that $\rho_i$'s do not differ from unit to unit. In the following we have considered such a case for large N.

For convenience in notation we shall supress the dependence of V on $\rho$ and denote $V(\rho)$ by V.

## 5.3 ESTIMATION

The parameters of the model considered in Section 5.2 may not always be identifiable. One such situation may arise when F is a general pxp matrix and in such a case parameters are not identifiable. We consider estimation of the parameters under the assumption that the parameters of the model are identifiable.

For the purpose of estimation of parameters of the model considered in the previous section, the ML method is adopted. It can be seen that the estimators of $\underset{\sim}{\mu}$, F, $\sigma^2$ and $\rho$ cannot be obtained by differentiating the likelihood as it is a very complicated function of the parameters. So the EM-algorithm (Dempster et al., 1977), which does not require the use of likelihood function, is used to obtain the ML estimates. The E- and M-steps of the algorithm are given below.

If $\underset{\sim}{\theta}$ were known, the expressions for estimators of the fixed and individual effects are well known (Harville, 1976) and are given by

$$\underset{\sim}{\tilde{\mu}} = (X' \ \Omega^{-1} \ X)^{-1} \ X' \ \Omega^{-1} \ \underset{\sim}{\tilde{y}} \ ; \qquad (5.3.1)$$

$$\underset{\sim}{\tilde{b}_i} = D(\underset{\sim}{\theta}) Z' \ \Omega^{-1} (\underset{\sim}{y_i} - X \underset{\sim}{\tilde{\mu}}) ; \qquad (5.3.2)$$

and their variances are

$$V_{\underset{\sim}{\tilde{\mu}}} = var(\underset{\sim}{\tilde{\mu}}) = N^{-1} (X' \ \Omega^{-1} \ X)^{-1} \ ; \qquad (5.3.3)$$

$$V_{\underset{\sim}{\tilde{b}_i}} = var(\underset{\sim}{\tilde{b}_i}) = D(\underset{\sim}{\theta}) Z' \ [\Omega^{-1} - \Omega^{-1} X\{var(\underset{\sim}{\tilde{\mu}})\} X'\Omega^{-1}] \ ZD(\underset{\sim}{\theta}). \quad (5.3.4)$$

The estimate of $\underset{\sim}{\mu}$ maximizes the likelihood based on the marginal distribution of the data. Of course, $\underset{\sim}{\tilde{b}_i}$ is not the maximum likelihood estimator but it can be derived by extension of Gauss-Markov theorem and is empirical Bayes (Harville, 1976), in the sense that it has the form

$$\underset{\sim}{\tilde{b}_i} = E(\underset{\sim}{\tilde{b}_i} | \underset{\sim}{y_i}, \underset{\sim}{\mu}, \underset{\sim}{\theta}).$$

If we were to observe $\underset{\sim}{b_i}$ (i.e. $\underset{\sim}{\beta_i}$ and $\underset{\sim}{e_i}$), we would find the maximum likelihood estimates of components of $\underset{\sim}{\theta}$, based on quadratic forms in $\underset{\sim}{b_i}$ as follows :

$$\underset{\sim}{\tilde{F}} = N^{-1} \Sigma \underset{\sim}{\beta_i} \underset{\sim}{\beta_i'} ; \qquad (5.3.5)$$

$\underset{\sim}{\tilde{\rho}}$ as a solution of the cubic

$$\rho^3 (p-1) \ R_0 - \rho^2 (p-2) R_1 - \rho \ [(p+1)R_0 + R_0'] + pR_1 = 0 \qquad (5.3.6)$$

which lies in $(-1, 1)$ (Anderson, 1971, p 353) and

$$\tilde{\sigma}^2 = (Np)^{-1}\left[(1+\tilde{\rho}^2)R_0 + R_0' - 2\tilde{\rho} R_1\right] \tag{5.3.7}$$

where

$$R_0 = \sum_2^p r_{ii} \ ; \ R_0' = r_{11}+r_{pp} \ ; \ R_1 = \sum_1^{p-1} r_{ii+1} \ ; \ R = \sum \tilde{e}_i \tilde{e}_i' = (r_{ij}).$$

Since $\tilde{b}_i$'s are not observable, we take the conditional expected values given $\tilde{y}_i$ and previous values of $\tilde{\mu}$ and $\tilde{\theta}$, say $\tilde{\mu}^{(k)}$ and $\tilde{\theta}^{(k)}$. It is not difficult to see that

$$E(\tilde{\beta}_i\tilde{\beta}_i' \mid \tilde{y}_i, \tilde{\mu}^{(k)}, \tilde{\theta}^{(k)}) = \tilde{\beta}_i^{(k)}\tilde{\beta}_i^{(k)'} + \text{var}(\tilde{\beta}_i \mid \tilde{y}_i, \tilde{\mu}^{(k)}, \tilde{\theta}^{(k)}) \ ;$$

$$E(\tilde{e}_i\tilde{e}_i' \mid \tilde{y}_i, \tilde{\mu}^{(k)}, \tilde{\theta}^{(k)}) = \tilde{e}_i^{(k)}\tilde{e}_i^{(k)'} + \text{var}(\tilde{e}_i \mid \tilde{y}_i, \tilde{\mu}^{(k)}, \tilde{\theta}^{(k)}) \ ;$$

where

$$\tilde{e}_i^{(k)} = \tilde{y}_i - X\tilde{\mu}^{(k)} - z_1\tilde{\beta}_i^{(k)}.$$

The equations (5.3.1) and (5.3.2) provide the E-step of the algorithm and the conditional expected values, given $\tilde{y}_i, \tilde{\mu}^{(k)}$ and $\tilde{\theta}^{(k)}$, of the quantities appearing in the equations (5.3.5) to (5.3.7) provide the M-step. Thus the algorithm is completely defined. These two steps are iterated till the convergence is achieved. Denote the converged values as $\hat{\tilde{\mu}}$ and $\hat{\tilde{\theta}}$ (i.e. $\hat{F}, \hat{\sigma}^2, \hat{\rho}$) which gives the maximum likelihood estimates.

As $\hat{\tilde{\theta}}$ is translation invariant and even function of the observations, $\hat{\tilde{\mu}}$ is unbiased for estimating $\tilde{\mu}$ (Kackar and Harville, 1984). Moreover, for the normal distribution considered here, the assumptions for ML estimates to have asymptotic optimum

properties are satisfied, and thus the estimators of $F$, $\sigma^2$ and $\rho$ are consistent and asymptotically efficient.

## 5.4 PREDICTION AND ITS MEAN SQUARE ERROR FOR A NEW INDIVIDUAL

Let us consider the problem of predicting the future value of the response of a new individual given the values of its past response $\underset{\sim}{y}_0$. To formalise the idea consider predicting the value of the response of the new individual at time $p+1$ ($y_{p+1,0}$, say), corresponding to the design points $\underset{\sim}{x}_2'$ and $\underset{\sim}{z}_2'$ belonging to the row spaces of $X$ and $Z_1$, respectively. It is not difficult to see that the minimum mean square error (MSE) predictor of $y_{p+1,0}$ given $\underset{\sim}{y}_0$, if $\underset{\sim}{\mu}$ and $\underset{\sim}{\theta}$ were known, is given by

$$Y_{p+1,0} = E(y_{p+1,0}|\underset{\sim}{y}_0)$$

$$= \underset{\sim}{x}_2' \underset{\sim}{\mu} + (\underset{\sim}{z}_2' F Z_1' + \sigma^2 \underset{\sim}{v}_{21}) Q^{-1}(\underset{\sim}{y}_0 - X\underset{\sim}{\mu})$$

$$= \underset{\sim}{x}_2' \underset{\sim}{\mu} + \underset{\sim}{z}_2' \underset{\sim}{\beta}_0 + \underset{\sim}{V}_{21} V_{11}^{-1} \underset{\sim}{e}_0$$

where

$$\underset{\sim}{V}_{21} = (1-\rho^2)^{-1} (\rho^p, \rho^{p-1}, \ldots, \rho).$$

However, if only $\underset{\sim}{\theta}$ were known, then it is natural to consider the predictor of $y_{p+1,0}$ as

$$\widetilde{Y}_{p+1,0} = \underset{\sim}{\lambda}' \underset{\sim}{\widetilde{\mu}} + \underset{\sim}{\delta}'(\underset{\sim}{\theta}) \underset{\sim}{\widetilde{b}}_0$$

where

$$\underset{\sim}{\lambda} = \underset{\sim}{x}_2 \ , \ \underset{\sim}{\delta}(\underset{\sim}{\theta}) = (\underset{\sim}{z}_2' : \underset{\sim}{V}_{21} V_{11}^{-1})'.$$

It is easily seen that $\underset{\sim}{V}_{21} V_{11}^{-1} = (0, 0, \ldots, \rho)$.

Let $M_1(\underset{\sim}{\theta})$ be the MSE of $\underset{\sim}{\tilde{y}}_{p+1,0}$. It can be seen that

$$M_1(\underset{\sim}{\theta}) = E(\tilde{y}_{p+1,0} - y_{p+1,0})^2$$

$$= \sigma^2 + \underset{\sim}{\lambda}' V_{\underset{\sim}{\mu}} \underset{\sim}{\lambda} + \underset{\sim}{\delta}'(\underset{\sim}{\theta}) V_{\underset{\sim}{b}_0 - \underset{\sim}{b}_0} \underset{\sim}{\delta}(\underset{\sim}{\theta}) - \frac{2}{N} \underset{\sim}{\lambda}' (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Z D(\underset{\sim}{\theta}) \underset{\sim}{\delta}(\underset{\sim}{\theta}),$$
$$(5.4.1)$$

where $V_{\underset{\sim}{b}_0 - \underset{\sim}{b}_0}$ is given by

$$V_{\underset{\sim}{b}_0 - \underset{\sim}{b}_0} = D(\underset{\sim}{\theta}) - V_{\underset{\sim}{b}_0}$$

and $V_{\underset{\sim}{b}_0}$ and $V_{\underset{\sim}{\mu}}$ are given in (5.3.3) and (5.3.4).

If all the parameters are unknown then it is natural to consider the predictor of $y_{p+1,0}$ as $\hat{y}_{p+1,0}$, given by,

$$\hat{y}_{p+1,0} = \underset{\sim}{\lambda}' \underset{\sim}{\hat{\mu}} + \underset{\sim}{\delta}'(\underset{\sim}{\hat{\theta}}) \underset{\sim}{\hat{b}}_0.$$

Thus there is an increase in MSE due to the use of estimated values in place of the parameters. Following Kackar and Harville (1984) it can be seen that $(\tilde{y}_{p+1,0} - y_{p+1,0})$ and $(\hat{y}_{p+1,0} - \tilde{y}_{p+1,0})$ are independent and hence the amount of increase in MSE is given by $E(\hat{y}_{p+1,0} - \tilde{y}_{p+1,0})^2$. If we expand $(\hat{y}_{p+1,0} - \tilde{y}_{p+1,0})^2$ by Taylor series in $\underset{\sim}{\hat{\theta}}$ about true values of parameters $\underset{\sim}{\theta}$, then by ignoring higher order terms it can be seen that

$$(\hat{y}_{p+1,0} - \tilde{y}_{p+1,0})^2 \simeq [\underset{\sim}{d}'(\underset{\sim}{\theta}) \cdot (\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta})]^2$$

where

$$\underset{\sim}{d}(\underset{\sim}{\theta}) = \partial \tilde{y}_{p+1,0} / \partial \underset{\sim}{\theta}.$$

Now

$$E(\hat{Y}_{p+1,0} - \tilde{Y}_{p+1,0})^2 = E\left[\underset{\sim}{d}'(\underset{\sim}{\theta})(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})\right]^2$$

$$= \text{Tr}\left[E\{\underset{\sim}{d}(\underset{\sim}{\theta})\underset{\sim}{d}'(\underset{\sim}{\theta})\} \cdot E\{(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})'\}\right] \qquad (5.4.2)$$

as $\underset{\sim}{d}(\underset{\sim}{\theta})$ and $\hat{\underset{\sim}{\theta}}$ are independent. Thus

$$E(\hat{Y}_{p+1,0} - \tilde{Y}_{p+1,0})^2 = \text{Tr}\left[A(\underset{\sim}{\theta}) \cdot B(\underset{\sim}{\theta})\right] \qquad (5.4.3)$$

where

$$A(\underset{\sim}{\theta}) = \text{var}\left[\underset{\sim}{d}(\underset{\sim}{\theta})\right] \; ; \; B(\underset{\sim}{\theta}) = E\left[(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})'\right].$$

It may be noted here that Kackar and Harville(1984) have used the expression similar to (5.4.2) in derivation of approximate MSE for predicting a future observation of the individual whose past observations have been used for estimation of $\underset{\sim}{\theta}$. This approximation may not be valid in general as it implies ignoring the terms of the same order as are being retained. They have discussed the conditions under which the expression (5.4.2) is exactly true.

The $(k,\ell)$-th element of the matrix $A(\underset{\sim}{\theta})$ can be obtained as (see Appendix 5.1)

$$a_{k\ell}(\underset{\sim}{\theta}) = N^{-1}\{\underset{\sim}{s}_k' \Omega \underset{\sim}{s}_\ell + N \underset{\sim}{t}_k'(\Omega + \Omega_1/N)\underset{\sim}{t}_\ell + 2\underset{\sim}{s}_k' \Omega_1 \underset{\sim}{t}_\ell\} \; ;$$

where $\underset{\sim}{s}'s$ and $\underset{\sim}{t}'s$ are as defined in the Appendix 5.1. The value of the matrix $B(\underset{\sim}{\theta})$ may be obtained as the inverse of information matrix. The terms necessary in the evaluation of the information matrix are given by

$$E\left(\frac{\partial^2 L}{\partial \theta_k \partial \theta_\ell}\right) = -\frac{N}{2} \operatorname{Tr}\{Z' \, \Omega^{-1} \, Z \cdot \frac{\partial D(\theta)}{\partial \theta_k} \, Z' \, \Omega^{-1} \, Z \cdot \frac{\partial D(\theta)}{\partial \theta_\ell}\} \cdot$$

From equations (5.4.1) and (5.4.3) we get that

$$E(\hat{y}_{p+1,0} - y_{p+1,0})^2$$

$$\cong \sigma^2 + \lambda' V_{\mu} \lambda + \delta'(\theta) V_{b_0 - b} \delta(\theta) - \frac{2}{N} \lambda' (X' \, \Omega^{-1} \, X)^{-1} X' \, \Omega^{-1} \, ZD(\theta)\delta(\theta)$$

$$+ \operatorname{Tr}\left[A(\theta) \cdot B(\theta)\right] \cdot \tag{5.4.4}$$

## 5.5 SOME SIMPLE ALTERNATIVE ESTIMATORS

In this section we consider two methods of obtaining initial estimates. For the sake of simplicity we consider $Z_1 = X$. These methods can be extended to a case when $Z_1 \neq X$ without much difficulty.

When $Z_1 = X$, the model (5.2.1) can be written as

$$y_i = X\alpha_i + e_i \, ,$$

where

$$\alpha_i = \mu + \beta_i \cdot$$

If $\rho$ (i.e. $V(\rho)$) is known then estimates of $F, \sigma^2$ and $\mu$ can be obtained as in Rao (1975) which are noniterative. In general, $\rho$ is unknown and has to be estimated from the data. Here, we suggest two methods of estimating $\rho$. These methods are iterative but involve numerical maximization of an objective function of one variable ($\rho$) only.

Method 1 : In this method $\rho$ is estimated by assuming $\underset{\sim}{\beta}_1$ (i.e. $\underset{\sim}{\alpha}_i$) to be fixed. If $\underset{\sim}{\alpha}_i$'s are fixed then the loglikelihood of $\underset{\sim}{y}_i$'s can be written as

$$\log L = -\frac{Np}{2} \log \sigma^2 - \frac{N}{2} \log|V| - \frac{1}{2\sigma^2} \text{Tr}\left[ V^{-1} \Sigma (\underset{\sim}{y}_i - X\underset{\sim}{\alpha}_i)(\underset{\sim}{y}_i - X\underset{\sim}{\alpha}_i)' \right].$$

$$(5.5.1)$$

The ML estimates of $\underset{\sim}{\alpha}_i$ and $\sigma^2$ are given by

$$\underset{\sim}{\alpha}_i^* = (X'V^{-1}X)^{-1}X'V^{-1}\underset{\sim}{y}_i \ ; \ \sigma_*^2 = \Sigma(\underset{\sim}{y}_i - X\underset{\sim}{\alpha}_i^*)' \ V^{-1}(\underset{\sim}{y}_i - X\underset{\sim}{\alpha}_i^*)/N(p-q)$$

The loglikelihood, after putting the values of $\underset{\sim}{\alpha}_i^*$'s and $\sigma_*^2$ in (5.5.1), is given by

$$\log L = \text{constant} - \frac{Np}{2}\log\sigma_*^2 - \frac{N}{2}\log|V|. \qquad (5.5.2)$$

The ML estimate is obtained by numerically maximizing (5.5.2). An alternative method is to differentiate (5.5.2) w.r.t. $\rho$ to obtain cubic equation similar to (5.3.6) and take an appropriate root, say $\rho^*$. Iterate between $\underset{\sim}{\alpha}_i^*, \sigma_*^2$ and $\rho^*$ till they converge.

This method is likely to give an underestimate of $\rho$ and the use of this value of $\rho$ may considerably affect estimation of other parameters.

The reason for getting an underestimate of $\rho$ is that it remains the same when $\underset{\sim}{\alpha}_i$'s are known or unknown. When p is small the substitution of $\underset{\sim}{\alpha}_i^*$ distorts the estimates of $\rho$ and $\sigma^2$ considerably, as seen by simulation. The effect of this distortion is small when p is very large. Moreover, when p is fixed the estimates of structural parameters $\rho$ and $\sigma^2$ obtained

in this way may not be consistent in presence of incidental parameters $\alpha_i$'s as discussed by Neyman and Scott (1948).

Method 2 : To avoid the difficulties involved in the method 1 we consider the estimates based on certain linear combinations of $y$'s whose expected values do not depend upon fixed effect parameters, as considered by Neyman and Scott (1948), Rao (1972) and Patterson and Thompson (1971). The proposed method is known as restricted maximum likelihood (REML). For details see Harville (1977).

Let W be any matrix that is orthogonal to X i.e. $W'X = O$, and define the vectors of residuals $\underset{\sim}{r}_i$'s as

$$\underset{\sim}{r}_i = W'\underset{\sim}{y}_i .$$

It follows that the first two moments of $\underset{\sim}{r}_i$'s are given by

$$E(\underset{\sim}{r}_i) = \underset{\sim}{O} ; \quad var(\underset{\sim}{r}_i) = \sigma^2 W'V(\rho)W,$$

where W is taken in such a way that $var(\underset{\sim}{r}_i)$ is non-singular. Estimate of $\rho$ and $\sigma^2$ can be obtained by maximizing the likelihood of $\underset{\sim}{r}_i$'s.

This method is free from the limitation of method 1 and can be expected to provide better estimate of $\rho$. It is not difficult to show that this method gives consistent estimator for large N.

When likelihood is based on $\underset{\sim}{r}_i$'s, instead of $\underset{\sim}{y}_i$'s, there are chances that absolute maximum of likelihood may occur outside the possible range or very close to the border.

Under these circumstances one may like to choose a value of $\rho$ in the neighbourhood of boundaries in the parameter space.

Once the estimator of $\rho$ (say, $\rho_*$) is obtained by the above methods, they can be used as though this were the true value of $\rho$ and estimates of $F$, $\sigma^2$ and $\underset{\sim}{\mu}$ may be obtained by using Rao's (1975) results as follows

$$N \underset{\sim}{\mu}_* = \Sigma \underset{\sim i}{\beta^\ell} \; ;$$

$$N(p-q)\sigma_*^2 = \Sigma (\underset{\sim i}{y} - X\underset{\sim i}{\beta^\ell})' (\underset{\sim i}{y} - X\underset{\sim i}{\beta^\ell}) \; ;$$

$$(N-1)(F+\sigma^2 U)_* = \Sigma (\underset{\sim i}{\beta^\ell} - \underset{\sim}{\mu}_*)(\underset{\sim i}{\beta^\ell} - \underset{\sim}{\mu}_*)'$$

where

$$\underset{\sim i}{\beta^\ell} = (X'V(\rho_*)^{-1}X)^{-1} X'V(\rho_*)^{-1}\underset{\sim i}{y}, \; U = (X'V(\rho_*)^{-1}X)^{-1}$$

When $\rho$ is known then the estimators of $\underset{\sim}{\mu}$, $F$ and $\sigma^2$ have the optimum properties discussed by Rao (1975) but when $\rho$ is not known and a consistent estimator is used then, in general, the above estimators provide consistent estimators of $F$ and $\sigma^2$. The estimator of $\underset{\sim}{\mu}$ is unbiased as well as consistent.

We shall call these two methods as empirical Bayes type (EB(1) and EB(2)) methods.

## 5.6  MONTE CARLO SIMULATION STUDY

To study the behaviour of the ML estimates and predictors on them, Monte Carlo simulations are carried out. The simulation study also enables the comparison of performance of ML method with methods EB(1) and EB(2). We can also examine the closeness

of expression for prediction MSE, given by (5.4.4), with MSD values. The data are generated in the following manner.

The data vectors of dimension 6 (i.e. p = 6) are generated. The design matrices X and $Z_1$ are taken to be the same and a linear (i.e. q = m = 2) fit is considered. The matrix X is taken as

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -3 & -1 & 1 & 3 & 5 \end{bmatrix} .$$

To study the performance of predictors obtained by various methods, the data corresponding to design point (1,7) are also generated. The vector of fixed effects $\underset{\sim}{\mu}$ is taken as

$$\underset{\sim}{\mu} = (4.00, \ 0.75)' .$$

The random vectors $\underset{\sim i}{\beta}$'s are generated from a bivariate normal distribution with mean $\underset{\sim}{0}$ and dispersion matrix F, given by

$$F = \begin{bmatrix} 2.00 & 0.00 \\ 0.00 & 1.40 \end{bmatrix} .$$

Five values of $\rho$(-0.40, 0.00, 0.15, 0.40 and 0.75) are considered. The value of $\sigma^2$ is chosen in such a way that the model explains, on an average, a predetermined percent of variability. Two levels (60% and 80%) of variability explained and two sample sizes (30 and 90) are considered. We restrict our study to only the selected combinations of these parameters.

For prediction purposes, we have also considered EB predictor that would have been obtained if we had considered

dispersion matrix of errors as $\sigma^2 I$ instead of $\sigma^2 V$.

As before, the random vectors are generated using the subroutine GGNRM of IMSL. One hundred runs are taken for each set of parameters. The summarized results are given in Tables 5.1 - 5.3.

## 5.7 RESULTS OF SIMULATION STUDY

### 5.7.1 Results on estimation of parameters :

From Tables 5.1 - 5.2 it is evident that the three methods (ML, EB(1) and EB(2)) perform similarly as for as $\underset{\sim}{\mu}$ is concerned. The estimates of $\rho$ by EB(1) is very highly underestimate. This affects estimation of other parameters also. EB(2) appears to provide better estimate of $\rho$ than EB(1). However, it can be seen that for $\rho = 0.40$ (and $\rho = 0.75$), $F(1,1)$ by EB(2) is very poorly estimated. The reason is that when the data are generated with $\rho = 0.40$ the estimated value of $\rho$ by EB(2) comes out to be very high ($> 0.90$). In such cases we take the value of $\rho_*$ as $0.90$. When the estimate of $\rho$ is high, the matrix $(X'V(\rho_*)^{-1}X)^{-1}$ is illconditioned (i.e. the ratio of the largest diagonal value to the smallest is very large). Since F is obtained by substituting $\sigma_*^2$ times this matrix from a positive definite matrix, F could fail to be positive definite. In our case, the $(1,1)$-th term of $(X'V(\rho_*)^{-1}X)^{-1}$ is very large when $\rho_* = 0.90$ and estimate of $F(1,1)$ is highly negative. Though there are few such cases they are dominating, affecting the average estimated value of $F(1,1)$. Because of this, the MSD value for this

Table 5.1:Estimates of Parameters $\mu$, F, $\sigma^2$ and $\rho$ and their Mean Square Deviations (MSD) by Methods of Maximum Likelihood(ML) and Empirical Bayes type(ER(1) and ER(2)) when 80% Variability is Explained by the Model

| Population Values | $\rho$ | Sample Size = 30 | | | | | | | Sample Size = 90 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ER(1) | ER(2) | ML | MSD(1) ×10 | MSD(2) ×10 | MSD(ML) ×10 | Asy.Var. ×10 | ER(2) | ML | MSD(2) ×10 | MSD(ML) ×10 | Asy.Var. ×10 |
| $\mu(1,1)$ =4.00 | -0.40 | 4.01 | 4.01 | 4.01 | 0.79 | 0.80 | 0.80 | 0.81 | 4.01 | 4.01 | 0.21 | 0.21 | 0.27 |
| | 0.00 | 4.01 | 4.01 | 4.01 | 0.88 | 0.92 | 0.91 | 0.92 | 4.02 | 4.02 | 0.33 | 0.32 | 0.42 |
| | 0.40 | 4.01 | 4.02 | 4.02 | 1.18 | 1.28 | 1.22 | 1.25 | | | | | |
| $\mu(2,1)$ =0.75 | -0.40 | 0.72 | 0.72 | 0.72 | 0.45 | 0.45 | 0.45 | 0.48 | 0.73 | 0.73 | 0.16 | 0.16 | 0.16 |
| | 0.00 | 0.72 | 0.72 | 0.72 | 0.46 | 0.46 | 0.46 | 0.49 | 0.73 | 0.73 | 0.17 | 0.17 | 0.17 |
| | 0.40 | 0.72 | 0.72 | 0.72 | 0.47 | 0.47 | 0.47 | 0.50 | | | | | |
| $F(1,1)$ =2.00 | -0.40 | 2.24 | 2.06 | 1.98 | 4.83 | 4.34 | 3.99 | 4.04 | 2.04 | 2.02 | 1.67 | 1.60 | 1.35 |
| | 0.00 | 2.54 | 1.97 | 1.92 | 8.60 | 6.95 | 5.88 | 5.82 | 0.57 | 1.99 | | 5.29 | 7.17 |
| | 0.40 | 3.59 | -1.51 | 1.97 | 37.0 | 822. | 10.5 | 21.5 | | 316. | 316. | | |
| $F(2,1)$ =0.00 | -0.40 | -0.01 | -0.01 | -0.01 | 1.47 | 1.46 | 1.36 | 1.17 | 0.02 | 0.02 | 0.33 | 0.32 | 0.39 |
| | 0.00 | -0.01 | -0.01 | -0.01 | 1.65 | 1.59 | 1.49 | 1.35 | 0.03 | 0.03 | 0.50 | 0.48 | 0.62 |
| | 0.40 | 0.01 | 0.01 | 0.01 | 2.36 | 2.15 | 2.00 | 1.87 | | | | | |
| $F(2,2)$ =1.40 | -0.40 | 1.40 | 1.38 | 1.34 | 1.48 | 1.47 | 1.41 | 1.39 | 1.39 | 1.38 | 0.45 | 0.44 | 0.46 |
| | 0.00 | 1.42 | 1.38 | 1.34 | 1.51 | 1.50 | 1.44 | 1.44 | 1.38 | 1.38 | 0.48 | 0.45 | 0.51 |
| | 0.40 | 1.46 | 1.37 | 1.34 | 1.62 | 1.68 | 1.50 | 1.52 | | | | | |
| $\sigma^2$ =4.58 | -0.40 | 4.00 | 4.66 | 4.65 | 6.84 | 4.97 | 4.93 | 4.66 | 4.63 | 4.62 | 1.83 | 1.72 | 1.55 |
| | 0.00 | 3.80 | 4.70 | 4.66 | 9.32 | 5.98 | 5.49 | 5.35 | 4.78 | 4.58 | 5.56 | 1.79 | 2.42 |
| | 0.40 | 3.45 | 4.91 | 4.52 | 15.7 | 13.1 | 5.36 | 7.26 | | | | | |
| $\rho$ | -0.40 | -0.64 | -0.38 | -0.38 | 0.62 | 0.11 | 0.11 | 0.10 | -0.39 | -0.40 | 0.03 | 0.03 | 0.03 |
| | 0.00 | -0.44 | 0.03 | 0.02 | 1.96 | 0.24 | 0.19 | 0.20 | 0.44 | 0.40 | 0.24 | 0.06 | 0.12 |
| | 0.40 | -0.28 | 0.47 | 0.38 | 4.69 | 0.60 | 0.20 | 0.36 | | | | | |

Table 5.2: Estimates of Parameters $\mu$, F, $\sigma^2$ and $\rho$ and their Mean Square Deviations (MSD) by

Methods of Maximum Likelihood(ML) and Empirical Bayes type(EB(2)) when 60%

Variability is Explained by the Model

| Population Values $\rho$ | Sample Size =30 | | | | Asy.Var. | Sample Size =90 | | | | Asy.Var. |
|---|---|---|---|---|---|---|---|---|---|---|
| | EB(2) | ML | MSD(2) | MSD(ML) | | EB(2) | ML | MSD(2) | MSD(ML) | |
| $\mu(1,1)$ -0.400 | 4.023 | 4.023 | 0.107 | 0.106 | 0.105 | 4.017 | 4.017 | 0.028 | 0.028 | 0.035 |
| =4.000   0.400 | 4.041 | 4.041 | 0.233 | 0.233 | 0.221 | 4.034 | 4.033 | 0.057 | 0.057 | 0.074 |
| $\mu(2,1)$ -0.400 | 0.729 | 0.729 | 0.049 | 0.047 | 0.051 | 0.730 | 0.730 | 0.016 | 0.016 | 0.017 |
| =0.750   0.400 | 0.724 | 0.725 | 0.053 | 0.053 | 0.056 | 0.733 | 0.733 | 0.019 | 0.019 | 0.019 |
| F(1,1) -0.400 | 1.994 | 1.928 | 0.739 | 0.647 | 0.732 | 2.028 | 2.005 | 0.276 | 0.260 | 0.244 |
| =2.000   0.400 | -6.167 | 2.656 | 465.7 | 2.859 | 10.745 | -1.962 | 2.450 | 221.7 | 1.280 | 3.582 |
| F(2,1) -0.400 | -0.021 | -0.020 | 0.198 | 0.184 | 0.160 | 0.016 | 0.015 | 0.049 | 0.048 | 0.053 |
| =0.000   0.400 | 0.007 | 0.002 | 0.435 | 0.392 | 0.369 | 0.039 | 0.039 | 0.101 | 0.097 | 0.123 |
| $\Gamma(2,2)$ -4.000 | 1.374 | 1.325 | 0.169 | 0.163 | 0.155 | 1.387 | 1.370 | 0.053 | 0.052 | 0.052 |
| =1.400   0.400 | 1.315 | 1.350 | 0.220 | 0.173 | 0.196 | 1.353 | 1.385 | 0.078 | 0.058 | 0.065 |
| $\sigma^2$ -0.400 | 12.482 | 12.397 | 3.698 | 3.489 | 3.300 | 12.355 | 12.325 | 1.287 | 1.239 | 1.100 |
| =12.222  0.400 | 13.090 | 11.767 | 8.665 | 3.213 | 4.887 | 12.768 | 12.002 | 3.862 | 1.033 | 1.629 |
| $\rho$ -0.400 | -0.374 | -0.380 | 0.012 | 0.010 | 0.009 | -0.387 | -0.389 | 0.003 | 0.003 | 0.003 |
| 0.400 | 0.464 | 0.350 | 0.056 | 0.015 | 0.032 | 0.445 | 0.378 | 0.024 | 0.004 | 0.011 |

Table 5.3: Observed and Theoretical Mean Square Error of Prediction

| N | $\sigma^2$ | $\rho$ | MSD ER(1) | MSD ER(2) | MSD ML | MSE (5.4.4) | MSE (5.4.1) | MSE ignoring $\rho$ |
|---|---|---|---|---|---|---|---|---|
| 30 | 4.583 | -0.40 | 9.471 | 8.862 | 8.862 | 8.970 | 8.839 | 10.261 |
| 30 | 4.583 | 0.00 | 9.805 | 8.267 | 8.241 | 8.383 | 8.212 | 8.149 |
| 30 | 4.583 | 0.15 | 9.701 | 7.969 | 7.922 | 8.094 | 7.909 | 7.987 |
| 30 | 4.583 | 0.40 | 9.610 | 7.356 | 7.286 | 7.496 | 7.297 | 8.036 |
| 30 | 4.583 | 0.75 | 9.450 | 6.273 | 6.188 | 6.391 | 6.204 | 8.038 |
| 30 | 12.222 | -0.40 |  | 22.918 | 22.902 | 23.158 | 22.711 | 26.635 |
| 30 | 12.222 | 0.40 |  | 18.919 | 18.584 | 19.094 | 18.466 | 20.590 |
| 90 | 4.583 | -0.40 |  | 8.945 | 8.940 | 8.877 | 8.834 | 10.290 |
| 90 | 4.583 | 0.40 |  | 7.438 | 7.370 | 7.352 | 7.286 | 8.189 |
| 90 | 12.222 | -0.40 |  | 23.023 | 23.013 | 22.822 | 22.675 | 26.575 |
| 90 | 12.222 | 0.40 |  | 18.903 | 18.651 | 18.622 | 18.416 | 20.862 |

estimator is also very large. The ML estimate of $\rho$ obtained by EM algorithm is satisfactory for all the cases and never gives these type of troubles.

In all the cases considered here, ML method gives estimators with smallest MSD; EB(1) gives estimators with highest MSD (except for F(1,1) when $\rho = 0.40$ where EB(2) gives very large MSD values for the reasons described above) and EB(2) generally gives better results than EB(1) but not as good as ML method.

It can be seen that the MSDs and asymptotic variances increase with increase in $\rho$ for all the estimators except for estimator of $\sigma^2$. This is expected because for $\rho$ near one the information contained in the observations about the parameters is small, whereas reverse is true when correlation is negative. The asymptotic variances of $\hat{F}$, $\hat{\sigma}^2$ and $\hat{\rho}$ are obtained from the inverse of information matrix. As sample size increases, the MSDs and asymptotic variances decrease, as one will expect.

The larger the amount of variability explained by the model, smaller are the MSDs and asymptotic variances. There is a better agreement between MSD and asymptotic variance when the amount of variability explained by the model is large than when it is small. These patterns are expected.

5.7.2 Results on prediction : Looking at the Table 5.3 for conditional prediction it can be seen that the ML produces MSD that

is smaller than by EB(1) and EB(2) methods. Thus the predictors
based on ML estimates give the best predictors though the
gain over EB(2) is not much. Thus for prediction purposes
EB(2) method can be chosen without much loss in efficiency.
In the last column of Table 5.3 we give MSD values of prediction
by ignoring the autocorrelation in prediction formula, though
they are actually present in the data. From the results
it is apparent that when the correlation is present in the
errors but prediction has been made by ignoring it then,
in general, there is a loss in efficiency. The loss is
substantial when $|\rho|$ is high but negligible when $\rho$ is moderate.

The formula (5.4.4) is applicable for a new individual
case but we have used it for predicting the future response
of an individual whose observations are used in the estimation
of parameters. Thus the MSDs and MSEs reported in Table 5.3
are not strictly comparable. Infact, the relevant MSD
corresponding to the value of MSE would be slightly higher
than the one reported and it has been confirmed by cross
validation studies. The expression for prediction MSE appears
to be in good agreement with MSD value and approximation (5.4.4)
is satisfactory.

As $\rho$ increases from -0.40 to 0.75, the MSD decreases
and so does its theoretical value obtain from equation (5.4.4).
For the values of X and F considered for simulation it can be
seen that the covariance with immediate successive observation

increases as $\rho$ increases. This means that the prediction would become more precise as is revealed by reduction in MSD and MSE values. However, this may not be true for all values of X and F.

Greater amount of variability explained by the model produces smaller MSD, as one will expect.

## 5.8 ILLUSTRATIVE EXAMPLE

To demonstrate the working of the method described in this chapter, a small example is worked out. The data of Williams and Izenman (1981) (see Experiment 2), which give weights of 13 (N=13) rats at the interval of 3 days from birth to weaning, are taken. The data are available for 7 time points but for estimation of the parameters data upto 6 (p=6) time points are used. The data at time point 7 are used for comparing the performance of the predictors.

The following model is considered.

$$\underset{\sim}{y}_i = X\underset{\sim}{\mu} + Z_1\underset{\sim}{\beta}_i + \underset{\sim}{e}_i \text{ , } i = 1,2,\ldots,13,$$

with $Z_1 = X$.

A quadratic fit is considered. Testing the goodness-of-fit of the model would be a significant contribution but for the timebeing that is not considered.

Estimates by method EB(2) : The estimates obtained by EB(2) come out to be

Table 5.4:Observed and Predicted Weights of Mice at t=7 by Maximum Likelihood(ML) and Empirical Bayes type(EB(2)) Methods

| Observed | Predicted Weights | |
|---|---|---|
| Weight | EB(2) | ML |
| 1.191 | 1.208 | 1.229 |
| 1.004 | 0.922 | 0.919 |
| 0.925 | 0.918 | 0.920 |
| 1.069 | 1.313 | 1.126 |
| 0.751 | 0.678 | 0.665 |
| 0.888 | 0.758 | 0.775 |
| 0.910 | 0.891 | 0.902 |
| 0.929 | 0.917 | 0.928 |
| 0.953 | 0.871 | 0.866 |
| 0.836 | 0.870 | 0.879 |
| 0.999 | 0.989 | 0.997 |
| 0.796 | 0.633 | 0.632 |
| 1.105 | 1.054 | 1.068 |
| MSD | $0.527 \times 10^{-2}$ | $0.537 \times 10^{-2}$ |

$$F_* = 10^{-3} \times \begin{bmatrix} -56.55 & & \\ 0.97 & 0.08 & \\ 0.09 & 0.00 & -0.01 \end{bmatrix} \; ; \; \sigma_*^2 = 0.216 \times 10^{-2} \; ; \; \rho_* = 0.983 \; ;$$

$$\underset{\sim}{\mu}_* = (0.583, \; 0.065, \; -0.009)'.$$

Estimates by ML method :    The estimates obtained by ML method come out to be

$$\hat{F} = 10^{-3} \times \begin{bmatrix} 4.99 & & \\ 0.96 & 0.19 & \\ -0.01 & 0.00 & 0.00 \end{bmatrix} \; ; \; \hat{\sigma}^2 = 0.104 \times 10^{-2} \; ; \; \hat{\rho} = 0.404 \; ;$$

$$\underset{\sim}{\hat{\mu}} = (0.583, \; 0.066, \; -0.009)'.$$

It can be seen that since the estimated value of $\rho$ by EB(2) is very high, thus the estimation of F and $\sigma^2$ is considerably affected for the reasons explained in Section 5.5.

Table 5.4 gives the observed and predicted weights of rats at t=7 by ML and EB(2) methods.  In this case the MSD by EB(2) method is slightly less than by ML method.  However, there is not much difference.  This is in agreement with the simulation results.  The MSD values are smaller than the minimum value (= .00643) reported by Rao (1981).  This is not surprising because in this model we have one more parameter.

5.9  DISCUSSION AND SUMMARY

Swamy (1971) considers a RCR model with errors following stationary autoregressive process of order 1 with each unit having its own autocorrelation coefficient.  His estimators are

satisfactory when there are large number of observations per
unit. Due to limitation of time and resources the observations
in longitudinal studies may hardly exceed eight or ten but it
is possible to choose N large. Analysis of RCR models when
errors have known dispersion matrix, except for constant
multiplier, is done by Rao (1975). In this chapter we consider
a mixed linear model with errors following stationary auto-
regressive process of order one when all the units have common
autocorrelation coefficient.

As in the previous chapter, for estimation of the
model parameters a ML estimation procedure via the EM algorithm
is adopted. Asymptotic variances of $\hat{F}, \hat{\sigma}^2$ and $\hat{\rho}$ are obtained
from the inverse of information matrix. Some alternative
simple estimators of empirical Bayes type are also considered.
Performances of these estimates and the ML estimates are studied
by performing Monte Carlo simulations.

Empirical Bayes type estimator constructed by Rao (1975)
have also been investigated for this problem. Two method
(EB(1) and EB(2)) for estimation of $\rho$ have been considered.
EB(1) is based on the likelihood of all observations assuming
regression parameters for each unit as fixed. This gives
considerable underestimate of $\rho$ for reasons explained in
Section 5.5. EB(2) is based on the likelihood of certain
linear combinations of observations free from fixed effects.
This method gives satisfactory results for $\rho$ not too large.

For positive and large $\rho$ this method sometimes gives very high value of $\rho_*$. In such cases the estimation of parameters is considerably affected. However, the ML procedure of Section 5.3 is satisfactory for all the cases. In all the cases considered here ML method gives estimators with least MSD; EB(1) generally gives estimators with highest MSD and EB(2) gives better results than EB(1) but not as good as ML method.

Conditional prediction on the basis of estimated values of the parameters obtained by the three methods is also studied. An expression for approximate prediction MSE, taking variability due to estimation of parameters, has been obtained in Section 5.4. The expression is complicated but is found to be satisfactory on the basis of simulation study. It is also seen that there is a better agreement in MSD and MSE as obtained from (5.4.4) when N is large, as one would expect. As for as prediction is concerned, ML method gives smaller MSD than by EB(1) and EB(2).

The simulation results are based on just 100 runs. This may not be adequate for a thorough simulation study. Moreover, the study is confined to few selected combinations of parameters and is not exhaustive. The purpose of study was to confirm some of the theoretical results and in most of the cases considered here there is a good agreement between simulation results and what we expect on theoretical grounds.

The method of estimation of the parameters can be modified without much difficulty when the errors follow nonstationary autoregressive process of order one. In that case the model of Mansour et al. (1985) can be treated as a particular case of this model.

APPENDIX 5.1

Derivation of $A(\underset{\sim}{\theta})$

$$A(\underset{\sim}{\theta}) = \text{var}\left[d(\underset{\sim}{\theta})\right] \; ; \; d(\underset{\sim}{\theta}) = \partial \tilde{y}_{p+1,0} / \partial \underset{\sim}{\theta}$$

where

$$\tilde{y}_{p+1,0} = \lambda' \underset{\sim}{\tilde{\mu}} + \delta'(\underset{\sim}{\theta}) \; \underset{\sim}{\tilde{b}}_0$$

and $\underset{\sim}{\tilde{\mu}}$ and $\underset{\sim}{\tilde{b}}_0$ are given by (5.3.1) and (5.3.2), respectively.

It is easy to observe that

$$\frac{\partial \underset{\sim}{\tilde{\mu}}}{\partial \theta_k} = -(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Z \frac{\partial D(\theta)}{\partial \theta_k} Z' P \underset{\sim}{\bar{y}} \qquad (i)$$

and

$$\frac{\partial \underset{\sim}{\tilde{b}}_0}{\partial \theta_k} = \left[I - D(\theta) Z' \Omega^{-1} Z\right] \frac{\partial D(\theta)}{\partial \theta_k} Z' \Omega^{-1} (\underset{\sim}{y}_0 - X \underset{\sim}{\tilde{\mu}})$$

$$+ D(\theta) Z' \Omega^{-1} X (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Z \frac{\partial D(\theta)}{\partial \theta_k} Z' P \underset{\sim}{\bar{y}} \qquad (ii)$$

where

$$P = \Omega^{-1} - \Omega^{-1} X (X' \Omega^{-1} X)^{-1} X' \Omega^{-1}.$$

From (i) and (ii) it follows that

$$d_k(\underset{\sim}{\theta}) = \partial \tilde{y}_{p+1,0} / \partial \theta_k$$

$$= \underset{\sim}{s}'_k \underset{\sim}{\bar{y}} + \underset{\sim}{t}'_k (\underset{\sim}{y}_0 - X \underset{\sim}{\tilde{\mu}})$$

where

$$\underset{\sim}{s}'_k = -\left[\underset{\sim}{\lambda}' - \underset{\sim}{\delta}'(\theta) D(\theta) Z' \Omega^{-1} X\right] (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Z \frac{\partial D(\theta)}{\partial \theta_k} Z' P$$

and

$$\underset{\sim}{t}'_k = \left[\underset{\sim}{\delta}'(\theta) \{I - D(\theta) Z' \Omega^{-1} Z\} \frac{\partial D(\theta)}{\partial \theta_k} + \frac{\partial \underset{\sim}{\delta}'(\theta)}{\partial \theta_k} D(\theta)\right] Z' \Omega^{-1}.$$

Noting that

$$\text{var}(\underset{\sim}{y}_0 - X\underset{\sim}{\tilde{\mu}}) = \{\Omega + \Omega_1/N\} \; ;$$

$$\text{cov}(\underset{\sim}{\bar{y}}, \underset{\sim}{y}_0 - X\underset{\sim}{\tilde{\mu}}) = \Omega_1/N$$

where

$$\Omega_1 = X(X'\Omega^{-1}X)^{-1} X'$$

it follows that $(k, \ell)$-th element of $A(\underset{\sim}{\theta})$ is given by

$$a_{k\ell}(\underset{\sim}{\theta}) = N^{-1}\{\underset{\sim}{s}'_k \Omega \underset{\sim}{s}_\ell + N\underset{\sim}{t}'_k(\Omega + \Omega_1/N)\underset{\sim}{t}_\ell + 2\underset{\sim}{s}'_k \Omega_1 \underset{\sim}{t}_\ell \} .$$

# DATA SETS

Experiment 1:The data of Strenio et al.(1983) are reported here.The data are modification of the data given in Box (1950) by adding random normal errors following $N(0,81)$ in the weight (in gm) of 10 rats of control group at a weekly interval.The mother's weight is artificially introduced as a covariate.

## Weights of rats in the control group

| Rat | Week | | | | | Mother's |
| | 0 | 1 | 2 | 3 | 4 | Weight |
|---|---|---|---|---|---|---|
| 1 | 61 | 72 | 118 | 130 | 176 | 170 |
| 2 | 65 | 85 | 129 | 148 | 174 | 194 |
| 3 | 57 | 68 | 130 | 143 | 201 | 187 |
| 4 | 46 | 74 | 116 | 124 | 157 | 156 |
| 5 | 47 | 85 | 103 | 117 | 148 | 155 |
| 6 | 43 | 58 | 109 | 133 | 152 | 150 |
| 7 | 53 | 62 | 82 | 112 | 156 | 138 |
| 8 | 72 | 96 | 117 | 129 | 154 | 154 |
| 9 | 53 | 54 | 87 | 120 | 138 | 149 |
| 10 | 72 | 98 | 114 | 144 | 177 | 167 |

Experiment 2:The weights of 13 male mice measured at succesive
intervals of 3 days from birth to weaning as reported by
Williams and Izenman(1981) are given in the following table.

Weight of 13 mice at intervals of 3 days from birth to weaning

| No./Days | 3 | 6 | 9 | 12 | 15 | 18 | 21 |
|---|---|---|---|---|---|---|---|
| 1 | 0.190 | 0.338 | 0.621 | 0.823 | 1.078 | 1.132 | 1.191 |
| 2 | 0.218 | 0.393 | 0.568 | 0.729 | 0.839 | 0.852 | 1.004 |
| 3 | 0.211 | 0.394 | 0.549 | 0.700 | 0.783 | 0.870 | 0.925 |
| 4 | 0.209 | 0.419 | 0.645 | 0.850 | 1.001 | 1.026 | 1.069 |
| 5 | 0.193 | 0.362 | 0.520 | 0.530 | 0.641 | 0.640 | 0.751 |
| 6 | 0.201 | 0.361 | 0.502 | 0.530 | 0.657 | 0.762 | 0.888 |
| 7 | 0.202 | 0.370 | 0.498 | 0.650 | 0.795 | 0.858 | 0.910 |
| 8 | 0.190 | 0.350 | 0.510 | 0.666 | 0.819 | 0.879 | 0.929 |
| 9 | 0.219 | 0.399 | 0.578 | 0.699 | 0.709 | 0.822 | 0.953 |
| 10 | 0.255 | 0.400 | 0.545 | 0.690 | 0.796 | 0.825 | 0.836 |
| 11 | 0.224 | 0.381 | 0.577 | 0.756 | 0.869 | 0.929 | 0.999 |
| 12 | 0.187 | 0.329 | 0.441 | 0.525 | 0.589 | 0.621 | 0.796 |
| 13 | 0.278 | 0.471 | 0.606 | 0.770 | 0.888 | 1.001 | 1.105 |

Experiment 3:The data of Elston and Grizzle(1962)
are reported here.The data consist of observations
on ramus height (in mm) for 20 boys of age 8 years
at four time points at half yearly intervals.

Ramus height data

|  | Age in years | | | |
| No. | 8.0 | 8.5 | 9.0 | 9.5 |
| --- | --- | --- | --- | --- |
| 1 | 47.8 | 48.8 | 49.0 | 49.7 |
| 2 | 46.4 | 47.3 | 47.7 | 48.4 |
| 3 | 46.3 | 46.8 | 47.8 | 48.5 |
| 4 | 45.1 | 45.3 | 46.1 | 47.2 |
| 5 | 47.6 | 48.5 | 48.9 | 49.3 |
| 6 | 52.5 | 53.2 | 53.3 | 53.7 |
| 7 | 51.2 | 53.0 | 54.3 | 54.5 |
| 8 | 49.8 | 50.0 | 50.3 | 52.7 |
| 9 | 48.1 | 50.8 | 52.3 | 54.4 |
| 10 | 45.0 | 47.0 | 47.3 | 48.3 |
| 11 | 51.2 | 51.4 | 51.6 | 51.9 |
| 12 | 48.5 | 49.2 | 53.0 | 55.5 |
| 13 | 52.1 | 52.8 | 53.7 | 55.0 |
| 14 | 48.2 | 48.9 | 49.3 | 49.8 |
| 15 | 49.6 | 50.4 | 51.2 | 51.8 |
| 16 | 50.7 | 51.7 | 52.7 | 53.3 |
| 17 | 47.2 | 47.7 | 48.4 | 49.5 |
| 18 | 53.3 | 54.6 | 55.1 | 55.3 |
| 19 | 46.2 | 47.5 | 48.1 | 48.4 |
| 20 | 46.3 | 47.6 | 51.3 | 51.8 |

Experiment 4:The data of a survey conducted by Indian Council of Agricultural Research on preharvest forecasting of yield of sugarcane in Meerut district of Uttar Pradesh in India during 1976-77 are reported here.The data consist of measurements on cane height (in m) and diameter (in cm) of randomly selected plants from 33 plots of area 25 sq.m. each. The data are taken at a monthly interval beginning from third month after sowing.The final yield (kg/plot) and number of shoots at time point 1 are also given.

Height and diameter of sugarcane

| Plot | Height (in m) | | | | | Diameter (in cm) | | | | | Yield/Plot in kg | No. of shoots |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | | |
| 1 | 0.84 | 1.48 | 1.75 | 2.06 | 2.18 | 8.60 | 7.90 | 8.40 | 8.40 | 8.00 | 26.38 | 61 |
| 2 | 0.53 | 1.48 | 1.78 | 1.92 | 2.05 | 8.20 | 8.30 | 7.60 | 7.70 | 7.90 | 36.40 | 110 |
| 3 | 0.87 | 1.13 | 1.25 | 1.85 | 2.01 | 7.40 | 8.10 | 8.20 | 7.60 | 7.20 | 74.50 | 193 |
| 4 | 0.66 | 0.74 | 1.14 | 1.18 | 1.25 | 6.70 | 6.80 | 6.40 | 6.10 | 5.90 | 31.30 | 66 |
| 5 | 0.74 | 0.84 | 1.13 | 1.16 | 1.21 | 7.30 | 7.50 | 6.90 | 6.50 | 6.30 | 29.55 | 48 |
| 6 | 0.94 | 1.04 | 1.38 | 1.61 | 1.68 | 7.90 | 8.40 | 6.70 | 6.40 | 6.30 | 51.40 | 120 |
| 7 | 0.89 | 0.99 | 1.34 | 1.71 | 1.75 | 8.40 | 8.70 | 6.70 | 6.70 | 6.70 | 47.95 | 99 |
| 8 | 1.04 | 1.31 | 1.55 | 1.79 | 1.81 | 7.10 | 7.80 | 6.70 | 6.60 | 6.50 | 41.00 | 77 |
| 9 | 1.14 | 1.30 | 1.65 | 1.90 | 1.92 | 6.50 | 8.50 | 6.50 | 6.30 | 6.30 | 43.00 | 74 |
| 10 | 0.47 | 1.00 | 1.28 | 1.50 | 1.60 | 7.30 | 7.50 | 7.40 | 7.30 | 7.10 | 30.00 | 93 |
| 11 | 0.52 | 0.98 | 1.30 | 1.44 | 1.49 | 7.10 | 7.50 | 7.10 | 6.50 | 6.30 | 28.00 | 63 |
| 12 | 0.29 | 0.88 | 1.10 | 1.40 | 1.49 | 7.10 | 6.80 | 6.70 | 6.70 | 6.70 | 47.00 | 140 |
| 13 | 0.49 | 1.50 | 1.85 | 1.91 | 1.98 | 7.90 | 7.80 | 7.50 | 7.30 | 7.10 | 44.00 | 100 |
| 14 | 0.65 | 1.18 | 1.51 | 1.80 | 1.87 | 7.62 | 7.82 | 7.58 | 7.62 | 7.50 | 55.00 | 125 |
| 15 | 0.77 | 0.95 | 1.30 | 1.50 | 1.55 | 7.79 | 7.22 | 7.20 | 6.79 | 6.82 | 50.00 | 120 |
| 16 | 0.52 | 1.06 | 1.49 | 1.89 | 1.91 | 8.00 | 6.90 | 6.96 | 6.78 | 6.54 | 52.00 | 95 |
| 17 | 0.66 | 0.73 | 0.82 | 0.92 | 0.95 | 6.84 | 6.70 | 6.78 | 6.72 | 6.74 | 31.00 | 70 |
| 18 | 1.10 | 1.58 | 1.66 | 1.72 | 1.87 | 6.56 | 6.94 | 7.32 | 6.96 | 6.88 | 35.25 | 115 |
| 19 | 0.60 | 0.84 | 0.99 | 1.12 | 1.26 | 7.14 | 6.70 | 6.94 | 6.76 | 6.80 | 32.50 | 110 |
| 20 | 0.82 | 1.55 | 1.72 | 1.81 | 1.87 | 6.56 | 6.36 | 6.24 | 6.22 | 6.16 | 41.25 | 115 |
| 21 | 0.52 | 0.98 | 1.46 | 1.74 | 1.87 | 6.76 | 6.50 | 6.48 | 6.62 | 6.48 | 43.00 | 113 |
| 22 | 1.40 | 1.51 | 1.70 | 1.88 | 1.89 | 8.04 | 6.70 | 6.64 | 6.74 | 6.82 | 73.00 | 190 |
| 23 | 1.46 | 1.50 | 1.70 | 1.83 | 2.06 | 8.16 | 6.82 | 6.62 | 6.48 | 7.04 | 69.00 | 192 |
| 24 | 1.07 | 1.30 | 1.58 | 1.75 | 1.78 | 7.88 | 6.70 | 6.64 | 6.62 | 6.64 | 60.00 | 180 |
| 25 | 0.94 | 1.25 | 1.59 | 1.82 | 1.84 | 7.66 | 6.20 | 6.10 | 6.14 | 6.08 | 47.50 | 95 |
| 26 | 0.84 | 1.36 | 1.65 | 1.77 | 1.80 | 8.20 | 8.70 | 8.20 | 7.90 | 7.52 | 52.00 | 135 |
| 27 | 0.71 | 1.34 | 1.39 | 1.43 | 1.48 | 7.90 | 8.50 | 7.64 | 7.50 | 7.40 | 51.00 | 117 |
| 28 | 0.99 | 1.51 | 1.63 | 1.90 | 2.02 | 8.42 | 8.40 | 7.62 | 7.42 | 7.06 | 50.00 | 154 |
| 29 | 1.00 | 1.57 | 1.67 | 1.92 | 2.01 | 7.40 | 7.20 | 7.12 | 6.72 | 6.14 | 59.00 | 138 |
| 30 | 0.99 | 1.14 | 1.22 | 1.41 | 1.60 | 6.96 | 6.88 | 6.98 | 6.90 | 6.74 | 55.00 | 146 |
| 31 | 0.92 | 1.30 | 1.39 | 1.74 | 1.87 | 7.14 | 6.88 | 7.08 | 7.04 | 6.92 | 50.00 | 143 |
| 32 | 0.80 | 1.01 | 1.31 | 1.70 | 1.96 | 8.46 | 6.86 | 6.76 | 6.64 | 6.52 | 43.00 | 137 |
| 33 | 0.79 | 1.03 | 1.09 | 1.40 | 1.73 | 9.14 | 6.58 | 6.22 | 6.32 | 6.24 | 39.00 | 128 |

## REFERENCES

Allen, D.M. (1971). Mean square error of prediction as a criterion
    for selecting variables. Technometrics, 13, 469-475.

Anderson, T.W. (1971). The statistical analysis of time series.
    John Wiley and Sons, New York.

Anderson, T.W. (1978). Repeated measurements on autoregressive
    processes. J. Amer. Statist. Assoc., 73, 371-378.

Bibby, J. and Toutenburg, H. (1977). Prediction and improved
    estimation in linear models. John Wiley and Sons,
    Toronto.

Box, G.E.P. (1950). Problems in the analysis of growth and wear
    curves. Biometrics, 6, 362-389.

Carter, R.L. and Yang, M.C.K. (1986). Large sample inference in
    random coefficient regression models. Commun. Statist.
    -Theor. Meth., 15, 2507-2525.

Chinchilli, V.M. and Carter, W.H. (1984). A likelihood ratio
    test for a patterned covariance matrix in a
    multivariate growth-curve model. Biometrics, 40,
    151-156.

Copas, J.B. (1983). Regression, prediction and shrinkage. J.
    Roy. Statist. Soc., B, 45, 311-354.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum
    likelihood from incomplete data via the EM
    algorithm. J. Roy. Statist.Soc., B, 39, 1-38.

Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K. (1981). Estimation
    in covariance components models. J. Amer.Statist.
    Assoc., 76, 341-353.

Diderrich, G.T. (1985). The Kalman filter from the perspective
    of Goldberger-Theil estimators. The American
    Statistician, 39, 193-198.

Dielman, T.E. (1983). Pooled cross-sectional and time series data:
    survey of current statistical methodology. The
    American Statistician, 37,111-122.

Duncan, D.B. and Horn, S.D.(1972). Linear dynamic recursive
    estimation from the viewpoint of regression analysis.
    J. Amer. Statist. Assoc., 67, 815-821.

Elston, R.C. and Grizzle, J.E. (1962). Estimation of time-response curves and their confidence bands. Biometrics, 18, 148-159.

Fearn, T. (1975). A Bayesian approach to growth curves. Biometrika, 62, 89-101.

Fearn, T. (1977). A two stage model for growth curves which leads to Rao's covariance adjusted estimators. Biometrika, 64, 141-143.

Fisk, P.R. (1967). Models of the second kind in regression analysis. J. Roy. Statist. Soc., B, 29, 266-281.

Geisser, S. (1970). Bayesian analysis of growth curves. Sankhya, A, 32, 53-64.

Geisser, S. (1980). Growth curve analysis. Handbook of Statistics, Vol. 1 (P.R. Krishnaiah, Ed.), 89-115. North-Holland Publishing Company, New York.

Geisser, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. Biometrika, 68, 243-250.

Gleser, L.J. and Olkin, I. (1966). A k-sample regression model with covariance. Multivariate Analysis - I (P.R. Krishnaiah, Ed.), 59-72. Academic Press, New York.

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalised linear regression model. J. Amer. Statist. Assoc., 57, 369-375.

Grizzle, J.E. and Allen, D.M. (1969). Analysis of growth and doz response curves. Biometrics, 25, 357-381.

Harrison, P.J. and Stevens, C.F. (1971). A Bayesian approach to short-term forecasting. Operations Research Quarterly, 22, 341-362.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting. J. Roy. Statist. Soc., B, 38, 205-247.

Harville, D.A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. Annals of Statistics, 4, 384-395.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc., 72, 320-340.

Hocking, R.R. (1974). Misspecification in regression. The American Statistician, 28, 39-40.

Hocking, R.R. (1976). The analysis and selection of variables in linear regression. Biometrics, 32,1-49.

Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Commun. Statist.-Theor. Meth., 10, 1249-1261.

Kackar, R N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J.Amer. Statist. Assoc., 79, 853-362.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. J. Basic Eng., 82, 34-45.

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curve. Ann. Inst. Statist. Math., 18, 75-86.

Khatri, C.G. (1973). Testing some covariance structures under a growth curve model. J. Multivariate Analysis, 3, 102-116.

Khatri, C.G. (1985). Some problems related to multivariate models, especially growth curve model. Proc. 72nd Ind. Sc. Cong. Part II, Lucknow, 1-20.

Kleinbaum,D.G. (1973). A generalisation of the growth curve model. J. Multivariate Analysis, 3,117-124.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 38,963-974.

Lee, J.C. and Geisser, S. (1972). Growth curve prediction. Sankhya, A,34,393-412.

Lee, J.C. and Geisser, S. (1975). Applications of growth curve prediction. Sankhya,A,37,239-256.

Lee, J.C. and Tan, W.Y. (1984). On the degree of polynomial in a general linear model, Commun. Statist. - Theor. Meth., 13, 781-790.

Lehmann, E.L. (1983). Theory of point estimation. John Wiley and Sons, New York.

Mansour, H., Nordheim, E.V. and Rutledge, J.J. (1985). Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors. Biometrics, 41, 287-294.

Maritz, M.S. (1970). Empirical Bayes methods. Methuen, London.

Meinhold, R.J. and Singpurwalla, N.D. (1983). Understanding the Kalman filter. The American Statistician, 37, 123-127.

Newbold, P. and Granger, C.W.J. (1974). Experience with forecasting univariate time series and the combination of forecasts. J. Roy. Statist. Soc., A, 137, 131-165.

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1-32.

Oza, M. and Shukla, G.K. (1986). Prediction under uncertainty of degree of polynomial in growth curve models. Commun. Statist. - Theor. Meth., 15, 1653-1664.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

Potthoff, R.F. and Roy, S.N. (1964). A generalised multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.

Rao, C.R. (1958). Some statistical methods for comparison of growth curves. Biometrics, 14, 1-17.

Rao, C.R. (1959). Some problems involving linear hypotheses in multivariate analysis. Biometrika, 46, 49-58.

Rao, C.R. (1961). Some observations on multivariate statistical methods in anthropological research. Bull. Inst. Statist. Math., 38, 99-111.

Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, 52, 447-458.

Rao, C.R. (1966). Covariance adjustment and related problems in multivariate analysis. Multivariate Analysis - I (P.R. Krishnaiah, Ed.), 87-103. Academic Press, New York.

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proc. Fifth Berkeley Symp., 1, 355-372.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. J.Amer. Statist. Assoc., 67, 112-115.

Rao, C.R. (1975). Simultaneous estimation of parameters in different linear models and applications to Biometric problems. Biometrics, 31, 545-554.

Rao, C.R. (1977). Prediction of future observations with special reference to linear models. Multivariate Analysis-IV (P.R.Krishnaiah, Ed.), 193-203. North-Holland Publishing Company, New York.

Rao, C.R. (1981). Prediction of future observations in polynomial growth curve models. Proc. of the Indian Statistical Institute Golden Jubilee International Conference on Statistics : Applications and New Directions, Calcutta, 512-520.

Rao, C.R. and Boudreau, R. (1985). Prediction of future observations in a factor analytic type growth model. Multivariate Analysis - IV (P.R.Krishnaiah, Ed.), 449-466. North-Holland Publishing Company, New York.

Rao, M.N. and Rao, C.R. (1966). Linked cross-sectional study for determining norms and growth rates - a pilot survey on Indian school-going boys. Sankhya, B, 28, 237-258.

Reinsel, G. (1982). Multivariate repeated measurement or growth curve models with multivariate random effects covariance structure. J. Amer. Statist. Assoc., 77, 190-195.

Reinsel, G. (1984a). Estimation and prediction in a multivariate random effects generalised linear model. J. Amer. Statist. Assoc., 79, 406-414.

Reinsel, G. (1984b). A note on conditional prediction in the multivariate linear model. J. Roy. Statist.Soc., B, 46, 109-117.

Reinsel, G. (1984c). Effects of the estimation of covariance matrix parameters in the generalised multivariate linear model. Commun. Statist.-Theor. Meth., 13, 639-650.

Reinsel, G. (1985). Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model. J. Amer. Statist. Assoc., 80, 642-650.

Rosenberg, B. (1973a). A survey of stochastic parameter regression. Annals of Economic and Social Measurement, 2, 381-397.

Rosenberg, B. (1973b). The analysis of a cross-section of time series by stochastically convergent parameter regression. Annals of Economic and Social Measurement, 2, 399-428.

Roy, S.N., Gnanadesikan, R. and Srivastava, J.N. (1971). Analysis and design of certain quantitative multiresponse experiments. Pergamon Press, Oxford.

Sallas, W.M. and Harville, D.A. (1981). Best linear recursive estimation for mixed linear models. J. Amer. Statist. Assoc., 76, 860-869.

Spjotvoll, E. (1977). Random coefficient regression models : A review. Math. Operationsforsch. Statist., Ser. Statistics, 8, 69-93.

Srivastava, J.N. (1966). Some generalisations of multivariate analysis of variance. Multivariate Analysis - I, (P.R. Krishnaiah, Ed.), 129-145. Academic Press, N.Y.

Srivastava, J.N. (1968). On a general class of designs for multiresponse experiments. Ann. Math. Statist., 39, 1825-1843.

Strenio, J.F. (1981). Empirical Bayes estimation for a heirarchical linear models. Unpublished Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Strenio, J.F., Weisberg, H.I. and Bryk, A.S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. Biometrics, 39, 71-86.

Swamy, P.A.V.B. (1971). Statistical inference in random coefficient regression models. Springer-Verlag, Berlin.

Swamy, P.A.V.B. and Mehta, J.S. (1977). Estimation of linear models with time and cross-sectionally varying coefficients. J. Amer. Statist. Assoc., 72, 890-898.

Swamy, P.A.V.B. and Tinsley, P.A. (1981). Linear prediction and estimation methods for regression models with stationary stochastic coefficients. Journal of Econometrics, 12, 103-142.

Timm,N.H. (1980). Multivariate analysis of variance of repeated measurements. Handbook of Statistics. Vol. 1 (P.R. Krishnaiah, Ed.), 41-37. North-Holland Publishing Company, New York.

Trawinski, I.M. and Bargmann, R.E. (1964). Maximum likelihood estimation with incomplete multivariate data. Ann. Math. Statist., 35, 647-657.

Verbyla, A.P. (1986). Conditioning in the growth curve model. Biometrika, 73, 475-483.

Ware, J.H. (1983). Growth curves. Encyclopedia of Statistical Sciences, Vol. 3 (S. Kotz and N.L. Johnson, Eds.), 539-542. John Wiley and Sons, New York.

Ware, J.H. and Berkey, C.S. (1984). Stochastic models for the effects of air pollutants on lung development. Proc. of the 12th Biometrics Conference, Tokyo, Japan, 217-227.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. The American Statistician, 39, 95-101.

Wegman, E.J. (1983). Kalman filtering. Encyclopedia of Statistical Sciences, Vol. 4 (S. Kotz and N.L. Johnson, Eds.), 342-346, John Wiley and Sons, New York.

West, M., Harrison, P.J. and Migon, S. (1985). Dynamic generalised linear models and Bayesian forecasting. J. Amer. Statist. Assoc., 80, 73-97.

Williams, J.B. and Izenman, A.J. (1981). A class of linear spectral models and analysis for the study of longitudinal data. Technical report, Department of Statistics, Colorado State University.

Wishart, J. (1938). Growth rate determination in nutrition studies with the bacon pig and their analysis. Biometrika, 30, 16-28.

Woolson, R.F., Leeper,J.D. and Clarke, W.R. (1978). Analysis of incomplete data from longitudinal and mixed longitudinal studies. J. Roy. Statist. Soc., A, 141, 242-252.

Woolson, R.F. and Leeper,J.D.(1980). Growth curve analysis of complete and incomplete longitudinal data. Commun. Statist.-Theor. Meth., 9, 1491-1513.

**Date Slip**

This book is to be returned on th
date last stamped.